INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(54) Title: MYCOBACTERIUM TUBERCULOSIS GENES ENCODING PROTEIN ANTIGENS

(57) Abstract

Mycobacterium tuberculosis genes encoding five immunologically relevant proteins have been isolated by systematically screening a lambda gt11 recombinant DNA expression library with a collection of murine monoclonal antibodies directed against protein antigens of this pathogen. One of the *M. tuberculosis* antigens, a 65kD protein, has been shown to have determinants common to *M. tuberculosis* and *M. leprae*. In addition, genes encoding proteins of other mycobacteria (*M. africanum, M. smegmatis, M. bovis* BCG and *M. avium*) have been isolated. Isolation and characterization of genes encoding major protein antigens of *M. tuberculosis* make it possible to develop reagents useful in the diagnosis, prevention and treatment of tuberculosis. They can be used, for example, in the development of skin tests, serodiagnostic tests and vaccines specific for tuberculosis.

AN3

-1-

MYCOBACTERIUM TUBERCULOSIS GENES AND
ENCODING PROTEIN ANTIGENS

Description

Background

05   Tuberculosis was the major cause of infectious
mortality in Europe and the United States in the
19th and early 20th centuries. Dubos, R. and J.
Dubos, The White Plague: Tuberculosis, Man and
Society, Little Brown & Co., Boston, MA, (1952).
10   Today, it remains a significant global health
problem.

        For example, in the United States there are
over 20,000 new cases of tuberculosis diagnosed
annually. In addition, the steadily declining
15   incidence of tuberculosis evident in preceding years
appears to have changed course, reaching a plateau
in 1985 and showing an increase in the first half of
1986. Centers for Disease Control, Morbidity/Mor-
tality, Weekly Report, 34:774 (1986); and Centers
20   for Disease Control, Morbidity/Mortality, Weekly
Report, 35:774 (1986).

        Worldwide, tuberculosis remains widespread and
constitutes a health problem of major proportions,
particularly in developing countries. The World
25   Health Organization estimates that there are ten
million new cases of active tuberculosis per year
and an annual mortality of approximately three

-2-

million. Joint International Union Against
Tuberculosis and World Health Organization Study
Group, Tubercle, 63:157-169 (1982).

Tuberculosis is caused by Mycobacterium (M.)
05   tuberculosis or Mycobacterium (M.) bovis, which are
the 'tubercle bacilli' of the family Mycobacteri-
aceae. M. bovis is a species which causes tubercu-
losis in cattle and is transmissible to humans and
other animals, in whom it causes tuberculosis. At
10   present, nearly all tuberculosis is caused by
respiratory infection with M. tuberculosis. Infec-
tion may be asymptomatic in some, but in other
individuals, it produces pulmonary lesions which
lead to severe debilitation or death. Resistance to
15   tuberculosis is provided by cell-mediated immune
mechanisms.

Mycobacteria are aerobic, acid-fast, non-spore-
forming, non-motile bacili with high lipid contents
and slow generation times. M. leprae is the eti-
20   ologic agent of leprosy and, among the other myco-
bacteria, the only major pathogen. Bloom, B.R. and
T. Godal, Review of Infectious Diseases, 5:765-780
(1983). However, other mycobacterial species are
capable of causing disease. Wallace, R.J. et.al.,
25   Review of Infectious Diseases, 5:657-679 (1984).
M.avium, for example, causes tuberculosis in fowl
and in other birds. Members of the M.
Avium-intracellularae complex have become important
pathogens among individuals with acquired immuno-
30   deficiency syndrome (AIDS). Certain groups of

individuals with AIDS have a markedly increased
incidence of tuberculosis-as well. Pitchenik, A.E.
et. al., Annals of Internal Medicine, 101:641-645
(1984).

05   Diagnostic and immunoprophylatic measures for
mycobacterial diseases have changed little in the
past half century. Tuberculin, developed by Koch as
a cure for tuberculosis in the late 1800s, is an M.
tuberculosis filtrate of complex and poorly-defined

10 composition. It is used as a skin test antigen to
detect prior exposure to the bacillus. Enrichment
of the protein fraction of this material in the
1930's produced the purified protein derivative
(PPD) which is still used to diagnose exposure to

15 tuberculosis. Seibert, F.M. et.al., American Review
of Tuberculosis, 30(Suppl.):705-778 (1934). Its
usefulness is limited, however, by its lack of
specificity and its inability to distinguish active
disease from prior sensitization by contact with M.

20 tuberculosis or cross-sensitization to other myco-
bacteria. Young, R.A. and R.W. Davis, Proceedings
of the National Academy of Sciences, USA, 80:194-
1198 (1983).

      Bacille Calmette Guerin (BCG), an avirulent

25 strain of M. bovis, has been used widely as a live
vaccine against tuberculosis for over 50 years.
Calmette, A., C. et.al., Bulletin of the Academy of

-4-

Medicine Paris, 91:787-796 (1924). During that
time, numerous studies have shown that BCG has
protective efficacy against tuberculosis. These
studies are reviewed by F. Luelmo in American Review
05   of Respiratory Diseases, 125(pt. 2):70-72 (1982).
However, more recently, a major trial of BCG in
India indicated that such a vaccine was not protec-
tive against tuberculosis in this setting. World
Health Organization WHO Technical Report Series, 651
10   (1980). Presently available approaches to diagnos-
ing, preventing and treating tuberculosis are
limited in their effectiveness and must be improved
if a solution is to be found for the important
public health problem tuberculosis represents
15   worldwide.

Summary of the Invention
        The present invention is based on the isolation
of genes encoding immunogenic protein antigens of
the tubercle bacillus Mycobacterium tuberculosis (M.
20   tuberculosis). Genes encoding such protein antigens
have been isolated from a recombinant DNA expression
library of M. tuberculosis DNA. Genes encoding
proteins of four additional mycobacteria have also
been isolated and restriction maps produced.
25   In particular, genes encoding five immunodomi-
nant protein antigens of the tuberculosis bacillus
(i.e., those M. tuberculosis proteins of molecular
weight 12,000 daltons (12kD), 14kD, 19kD, 65kD and
71kD have been isolated by probing a lambda gt11
30   expression library of M. tuberculosis DNA with

monoclonal antibodies directed against M. tuberculosis-specific antigens.

Recombinant DNA clones producing the specific antigenic determinants recognized by the monoclonal antigens were also isolated in this manner. DNA from such recombinant lambda gt11 clones was mapped with restriction endonucleases; the restriction maps for genes encoding the five immunodominant protein antigens (i.e., genes encoding the 12kD, 14kD, 19kD, 65kD and 71kD proteins) were deduced. The nucleotide sequence of three of the genes have been determined and, in each case, the amino acid sequence of the encoded protein has been deduced.

Brief Description of the Drawings

Figure 1 shows restriction maps of M. tuberculosis DNA. Recombinant DNA clones isolated with monoclonal antibodies directed against the 12kD, 14kD, 19kD, 65kD and 71kD protein antigens were mapped with restriction endonucleases. The insert DNA endpoints are designated left (L) or right (R) in relation to lac Z transcripts which traverse the insert from right to left. Restriction sites are represented as follows: A, Sal I; B, BamHI; E, EcoRI; G, BglII; K, KpnI; P, PvuI; S, SacI; X, XhoI.

Figure 2 shows arrays of antigens from M. tuberculosis recombinant DNA clones probed with rabbit hyperimmune serum. The code of the recombinant DNA clones shown on the numbers filter is: 1, Y3275; 2, Y3274; 3, Y3279; 4, Y3277; 5, Y3247; 6, Y3272; 7, Y3150; 8, Y3254; 9, Y3147; 10, Y3163; 11,

-6-

Y3179; 12, Y3191; 13, Y3252; 14, Y3178; 15, Y3180;
16, Y3143; 17, lambda gt11. Clones 1, 5, 6, 7, 9
and 16 are M. tuberculosis recombinants described in
the following section. Clones 10, 11, 14 and 15 are
05   M. leprae recombinants expressing epitopes of the
18kD, 28kD, 36kD and 65kD antigens, respectively.
Clones 2, 3, 4, 8, 12, 13 are uncharacterized
recombinants from the lambda gt11 M. tuberculosis
and M. leprae libraries. Clone 17 is a non-
10   recombinant lambda gt11 control.

Figure 3 shows arrays of recombinant mycobac-
terial antigens probed with monoclonal antibodies to
assess the extent of cross-reactivity between
recombinant protein antigen of M. tuberculosis and
15   of M. leprae. The array of clones is identical to
that shown in Figure 2. Antibody probes and the
antigen sizes recognized are: 1, IT-11 (71kD); 2,
IT-31 (65kD); 3, IT-16 (19kD); 4, IT-1 (14kD); 5,
IT-3 (12kD).
20   Figure 4 shows restriction maps of DNA encoding
four proteins (71kD, 65kD, 19kD and 14kD) of M.
tuberculosis and four proteins (71kD, 65kD, 19kD and
14kD) of M. bovis BCG. Restriction sites are
represented as follows: A, AatII; B, BamHl; C,
25   BclI; D, DraIII; E=EcoRI; G, BglII; H, HinfI; K,
KpnI; P, PstI; S, SalI; V, PvuI and X, XhoI.
Figure 5 is a comparison of restriction maps of
the gene encoding the 65kD protein of 6 mycobacteria
(M. leprae, M. tuberculosis, M. africanum, M. bovis
30   BCG, M. smegmatis, M. avium). Restriction sites are

-7-

as follows: B, BamHl; K, KpnI; N, SacI; P, PvuI; S, SalI; X, XhoI.

    Figure 6 is the nucleotide sequence of the region containing the M. tuberculosis 19kD gene. The deduced amino acid sequence of the encoded protein is also represented (protein start position, nucleotide 1110; protein stop position, nucleotide 1586).

    Figure 7 is the nucleotide sequence of the region containing the M. tuberculosis 71kD gene and the deduced amino acid sequence of the encoded protein.

    Figure 8 is the nucleotide sequence of the region containing the M. tuberculosis 65kD gene. The deduced amino acid sequences of the two long open reading frames are presented in one letter code over (540) or under (517) the appropriate triplets.

Detailed Description of the Invention

    The invention described herein is based on the isolation of genes encoding immunogenic protein antigens of the bacillus M. tuberculosis, which is the major etiologic agent of tuberculosis. In particular, it is based on the isolation, using monoclonal antibodies directed against M. tuberculosis-specific antigens, of genes encoding five immunogenic protein antigens of the tuberculosis bacillus; these five antigens are immunodominant. Immunogenic antigens are those which elicit a response from the immune system. Immunodominant protein antigens are immunogenic

-8-

antigens against which the immune system directs a
significant portion of its response. Genes encoding
M. tuberculosis antigens of molecular weight 12,000
daltons (12kD), 14kD, 19kD, 65kD and 71kD were
05　　isolated in this manner.

Isolation and characterization of major protein
antigens of M. tuberculosis, as described herein,
make it possible to develop more effective tools for
the prevention, diagnosis, and treatment of tubercu-
10　　losis. Identification and isolation of genes
encoding five immunodominant M. tuberculosis protein
antigens, as well as of the five protein antigens,
are described below; uses of the genes and encoded
products are also described.

15　　　　M. bovis BCG DNA clones were also isolated for
the genes encoding the 71kD, 65kD, 19kD and 14kD
proteins. In order to compare M. bovis BCG and M.
tuberculosis genes encoding proteins of similar
molecular weight, restriction endonuclease maps were
20　　determined for DNA segments containing each of the
genes. Restriction maps for each of these genes is
represented in Figure 4.

In addition, DNA clones were isolated for the
genes encoding the 65kD protein from M. africanum,
25　　M. smegmatis and M. avium. Restriction endonuclease
maps were determined for DNA segments containing
each of these genes. The restriction maps for these
genes, as well as for the genes encoding the 65kD
protein of M. tuberculosis, M. bovis BCG and M.
30　　leprae, are represented in Figure 5.

-9-

I.   Construction of a recombinant expression
      library of M. tuberculosis DNA

A recombinant DNA expression library of M.
tuberculosis DNA was constructed using lambda gt11.
05   The library was constructed with M. tuberculosis
genomic DNA fragments in such a way that all
protein-coding sequences would be represented and
expressed. Young, R.A., B.R. Bloom, C.M.
Grosskinsky, J. Ivanyi, D. Thomas and R.W. Davis,
10   Proceedings of the National Academy of Sciences,
USA, 82:2583-2587 (1985).

Lambda gt11 is a bacteriophage vector which is
capable of driving the expression of foreign insert
DNA with E. coli transcription and translation
15   signals. Lambda gt11 expresses the insert DNA as a
fusion protein connected to the E. coli Beta-
galactosidase polypeptide. This approach ensures
that the foreign DNA sequence will be efficiently
transcribed and translated in E. coli. This ap-
20   proach is also useful in addressing the problem of
the highly unstable nature of most foreign proteins;
fusion proteins are often more resistant to prote-
olytic degradation than is the foreign polypeptide
alone. Lambda gt11 and the E. coli strain used
25   (Y1090) have been described previously. Young, R.A.
et al., Proceedings of the National Academy of
Sciences, USA, 80:1194-1198 (1983); Young, R.A. and
R.W. Davis, Science, 222:778-782 (1983). The
teachings of these publications are incorporated
30   herein by reference. The library constructed in
this manner has a titer of 1x 10^{10} pfu/ml. and

-10-

contains approximately 40% recombinants with an
average insert size of 4kB.

II. Screening of the lambda gt11 <u>M. tubercu-</u>
<u>losis</u> library with antibody probes

05   Murine monoclonal antibodies to protein anti-
gens of <u>M</u>. tuberculosis were used individually to
probe the <u>M. tuberculosis</u> recombinant DNA library.
This work is described below and with specific
reference to the 65kD antigen in the Exemplifica-
10   tion. The antibodies used as probes and the sizes
of the antigens to which they bind are shown below.

	<u>Antibody</u>	<u>M. tuberculosis</u> <u>Antigen</u>
	IT-3	12kD
15	IT-20	14kD
	IT-19	19kD
	IT-27	19kD
	IT-17	23kD
	IT-29	23kD
20	IT-15	38kD
	IT-21	38kD
	IT-23	38kD
	IT-13	65kD
	IT-31	65kD
25	IT-33	65kD
	IT-11	71kD

Engers, H.D. <u>et al., Infectious Immunology,</u>
<u>51</u>:718-720 (1986).

-11-

All monoclonal antibodies were used at approximately
1:200 to 1:300 dilution in 50mM Tris-HCl pH8/150 mM
NaCl/.05% Tween 20.

05   Screening of the lambda gt11 recombinant DNA
library was performed as described by Young et al.
in Proceedings of the National Academy of Sciences,
USA, 82:2583-2587 (1985), the teachings of which are
incorporated herein by reference. One modification
was made in the method described by Young and

10   co-workers: 1% bovine serum albumin was used in
place of 20% fetal calf serum to decrease back-
ground.

Briefly, cloned lambda gt11 recombinants were
arrayed on lawns of E. coli Y1090. The phage were

15   grown, antigen expression was induced and the
antigens were blotted and probed with serum.
Detection of signal-producing plaques was performed
with a biotinylated secondary antibody system
(Vectastain, Vector Laboratories, Burlingame, CA) or

20   with an alkaline phosphatase conjugated secondary
antibody system (Protoblot, Promega Biotec, Madison,
WI), both used according to manufacturer's instruc-
tions. Signal-producing clones were isolated using
antibodies directed against protein antigens of

25   molecular weight 12kD, 14kD, 19kD and 65kD and 71kD.
In each case, similar numbers of clones were
isolated in screens of approximately 10^5 recombinant
plaques. DNA clones encoding the 23kD and 38kD
antigens could not be detected with these anti-

30   bodies, possibly because the native epitope is
modified or topographically complex, or because the

-12-

antigen-antibody interaction is too weak to be
recognized by current detection methods.

III. Probing of Arrays of lambda gt11 DNA
Clones with Antibody Probes

05   0.2 ml of a saturated culture of Y1090 was
added to 2.5 ml of molten LB soft agar, poured onto
100 mm plates containing 1.5% LB agar and allowed to
harden at room temperature for 10 min. 100 ul of
phage plate stock containing approximately 10^{11}
10   pfu/ml of the lambda gt11 DNA clones of interest
were placed into alternate wells of 96-well tissue
culture plates. A multi-pronged transfer device was
placed briefly in the wells containing phage and
then touched lightly to the surface of the plate
15   onto which the soft agar had been poured. The
plates were then incubated at 42°C for approximately
3 hours, at which point clear plaques approximately
5mm in diameter were visible. The plates were then
overlayed with nitrocellulose filters saturated with
20   10mM isopropylthiogalactoside (IPTG) and incubated
at 37°C for 3.5 hours. Subsequent processing of
filters for detection of antigen was identical to
the procedures described for screening of lambda
gt11 library with antibody probes.
25   Immunoscreening of the lambda gt11 library to
isolate clones reactive with monoclonal antibodies
specific for the 65kD antigen is described in the
Exemplification.

-13-

IV. Recombinant DNA Manipulation

DNA from recombinant lambda gt11 clones was
isolated and mapped with restriction endonucleases
by standard techniques. Davis, R.W. et al.,
05   Advanced Bacterial Genetics: A Manual for Genetic
Engineering, Cold Spring Harbor (1980).

Figure 1 shows the genomic DNA restriction map
deduced for each of the genes encoding the five M.
tuberculosis antigens and illustrates how each of
10   the cloned DNAs aligns with that map. All clones
isolated with monoclonal antibodies directed against
any single antigen align with a single genomic DNA
segment. This indicates that all clones were
isolated because they express the protein of inter-
15   est rather than an unrelated polypeptide containing
a similar or identical epitope. In addition, this
result suggests that each antigen is the product of
a single gene.

The orientation of each DNA insert in the
20   recombinant clones was determined by restriction
analysis. Only among the clones for the 65kD
antigen were the inserts found in both possible
orientations relative to the direction of lac Z
transcription in lambda gt11. This suggests that
25   this protein can be expressed in E. coli from
signals independent of those provided by lac Z.
Similar results have been obtained for recombinant
DNA clones encoding the 65kD antigens of M. bovis
and M. leprae. Thole, J.E.R. et al., Infectious
30   Immunology, 50:800-806 (1985); Young, R.A. et al.,
Nature, 316:450-452 (1985).

-14-

The nucleotide sequences of three regions of
the M. tuberculosis DNA were determined: 1) the
region containing the M. tuberculosis 19kD gene; 2)
the region containing the M. tuberculosis 71kD gene;
05   and 3) the region containing the 65kD gene. The
three sequences are represented in Figures 6-8.
Sequences were determined using standard techniques,
which are described in the Exemplification.

       V.   Filter hybridization of Insert DNA
10   Arrays of lambda gt11 clones were created as
described above and incubated at 42° for 5 hours.
The plates were then overlayed with nitrocellulose
filters and placed at 4°C for 1 hour. Probe DNA was
labelled with ^{32}P by nick translation. Filter
15   hybridization was performed as described by Davis et
al. in Advanced Bacterial Genetics: A Manual for
Genetic Engineering, Cold Spring Harbor (1980), the
teachings of which are incorporated herein by refer-
ence. Hybridization conditions were as follows: 50%
20   v/v formamide, 5x SSPE (1x SSPE is 0.18M NaCl, 10mM
$Na_{1.5}H_{1.5}PO_4$, 1mM Na_2 EDTA, pH 7.0), 1x Denhardt's
solution (0.02% w/v Ficoll, 0.02% w/v polyvinyl-
pyrrolidone, 0.02% w/v bovine serum albumin), 0.3%
$NaDodSO_4$ at 42°C for approximately 16 hours, fol-
25   lowed by washing in 2x SSPE, 0.2% $NaDodSO_4$ at 45°C.

       VI.   Recombinant Antigens Recognized by Rabbit
             Serum
       The response of a second animal to an antigen
preparation of M. tuberculosis was assessed by

-15-

examining the reactivity of rabbit anti-M. tuberculosis hyperimmune sera with recombinant antigens. Cloned lambda gt11 recombinants were arrayed on lawns of E. coli and probed with the rabbit hyper-
05   immune serum. Anti-M. tuberculosis hyperimmune serum, produced by repeated immunization of rabbits with M. tuberculosis H37Rv culture filtrate, was provided by J. Bennedsen (Statens Seruminstitut, Copenhagen, Denmark). These sera were used at 1:100
10   dilution.

        These sera produced positive signals with lambda gt11 clones encoding each of the five M. tuberculosis epitopes which had been isolated with murine monoclonal antibodies (Figure 2). Particu-
15   larly strong signals were observed with the 65kD and 71kD antigens (Figure 2). These results demonstrate that mice and rabbits can mount an antibody response to the same protein antigens of M. tuberculosis.

        Clones for the five M. tuberculosis antigens
20   were detected at similar frequencies in the lambda gt11 recombinant DNA library. Thus, the number and type of antigen-producing clones isolated with polyclonal serum antibodies should reflect the relative titer and deversity of the individual
25   antibodies in this serum.

        To determine whether any of the 5 M. tuberculosis antigens are relatively immunodominant in the rabbit humoral immune response to M. tuberculosis, the M. tuberculosis lambda gt11 recombinant DNA
30   library was screened with the rabbit serum. Forty signal-producing clones were isolated, arrayed on

-16-

lawns of E. coli Y1090 and probed with monoclonal
antibodies directed against each of the 5 recombi-
nant M. tuberculosis protein antigens. Remarkably,
17 of the 40 clones (43%) reacted strongly with at

05   least one of the four anti-65kD monoclonal anti-
bodies tested. An additional six clones (15%)
reacted strongly with the anti-17kD monoclonal
antibody (IT-11). This indicates that a large
proportion of the anti-M. tuberculosis antibody

10   present in the rabbit serum was directed against the
65kD antigen of M. tuberculosis and suggests that it
is a dominant antigen for the rabbit humoral immune
response. Seventeen of the clones did not react
with any of the monoclonal antibodies tested,

15   suggesting that the rabbit sera may identify M.
tuberculosis proteins not recognized by the murine
antibodies.

      VII. Antigenic Relatedness of M. tuberculosis
           and M. leprae Proteins

20   There is evidence that M. tuberculosis and M.
leprae share immunologically important antigens. To
assess this further, an investigation of the exact
nature of the immunological relatedness among
recombinant protein antigens of M. tuberculosis and

25   M. leprae was conducted.
          For each of five M. tuberculosis and four M.
leprae protein antigens, a single recombinant DNA
clone containing most or all of the gene of interest
was used to express antigen in the following manner.

30   The recombinant phage clones were arrayed on a lawn

-17-

of <u>E</u>. <u>coli</u> Y1090, which was then grown and induced
for antigen expression. ._

    Antigen immobilized on nitrocellulose filters
was then probed with 26 individual anti-<u>M</u>. <u>tubercu-</u>
05   <u>losis</u> and <u>M</u>. <u>leprae</u> monoclonal antibodies. Figure 3
shows the array of DNA clones used and the results
obtained with the anti-M <u>tuberculosis</u> antibodies
IT-1, IT-3, IT-11, IT-16, and IT-31, which recog-
nized proteins of 14kD, 12kD, 71kD, 19kD and 65kD
10   respectively. Table 1 details the full results of
these cross-screening experiments, showing the
reactivity of antigen expressed from individual
recombinant DNA clones with each of the individual
monoclonal antibodies. Clones were scored as
15   positive only if the signal produced was clearly
greater than the background signal produced by the
non-recombinant lambda gt11 clone included in each
array.

-18-

TABLE 1

Reactivity of Monoclonal Antibodies with Recombinant Protein Antigens

		DNA CLONES									
		M. tuberculosis						M. leprae			
		12kD	14kD	19kD	65kD	71kD	-	18kD	28kD	36kD	65kD
ANTIBODIES		Y3275	Y3247	Y3147	Y3150	Y3272	λgt11	Y3179	Y3163	Y3180	Y3178
M. tuberculosis											
12kD	IT-3	⊙	-	-	-	-	-	-	-	-	-
14kD	IT-1	-	⊙	-	-	-	-	-	-	-	-
	IT-4	-	⊙	-	-	-	-	-	-	-	-
	IT-20	-	⊙	-	-	-	-	-	-	-	-
19kD	IT-10	-	-	⊙	-	-	-	-	-	-	-
	IT-12	-	-	⊙	-	-	-	-	-	-	-
	IT-16	-	-	⊙	-	-	-	-	-	-	-
	IT-19	-	-	⊙	-	-	-	-	-	-	-
65kD	IT-13	-	-	-	⊙	-	-	-	-	-	-
	IT-31	-	-	-	⊙	-	-	-	-	-	⊙
	IT-33	-	-	-	⊙	-	-	-	-	-	⊙
71kD	IT-11	-	-	-	-	⊙	-	-	-	-	-
M. leprae											
18kD	L7-15	-	-	-	-	-	-	⊙	-	-	-
28kD	SA1.D2D	-	-	-	-	-	-	-	⊙	-	-
	SA1.BIIH	-	-	-	-	-	-	-	-	-	-
36kD	F47-9-1	-	-	-	-	-	-	-	-	⊙	-
	HLO4-A	-	-	-	-	-	-	-	-	-	-
65kD	CI.I	-	-	-	⊙	-	-	-	-	-	⊙
	IIH9	-	-	-	⊙	-	-	-	-	-	⊙
	IIIE9	-	-	-	-	-	-	-	-	-	⊙
	IIC8	-	-	-	⊙	-	-	-	-	-	⊙
	IIIC8	-	-	-	-	-	-	-	-	-	⊙
	T2.3	-	-	-	⊙	-	-	-	-	-	⊙
	Y1-2	-	-	-	⊙	-	-	-	-	-	⊙
	SA2.D7C	-	-	-	⊙	-	-	-	-	-	⊙
	HL30A	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙

-19-

Several conclusions can be drawn from the
results shown in Table 1. Among the 11 monoclonal
antibodies that recognize a 65kD antigen, 7 react
with the 65kD protein from both mycobacteria (IT-31,
05   C1.1, IIH9 (identical to IT-33), IIC8, T2.3, Y1-2,
SA2.D7C), one antibody reacts only with the M.
tuberculosis 65kD protein (IT-13), and two
antibodies react only with the M. leprae 65kD
protein (IIIE9 and IIIC8). One antibody, ML30A,
10   cross-reacts with an antigen in E. coli and could
not specifically identify antigen-producing clones.
These results indicate that the 65kD protein
antigens of M. tuberculosis and M. leprae are
homologues and share a number of epitopes. In
15   addition to these shared epitopes, however, both
65kD antigens have epitopes that are specific for
one species relative to the other.
        No cross-reactivity was observed between other
antigens of these two mycobacterial species.
20   Because monoclonal antibodies recognize a single
epitope and because only one or a few antibodies
were available for each antigen, it is not clear
whether the 65kD proteins are the only homologous
protein antigens of M. tuberculosis and M. leprae.
25   Among the antigens for which lambda gt11 clones have
been isolated, the 18kD antigen of M. leprae and the
19kD antigen of M. tuberculosis are of similar size.
To determine whether these two antigens are related,
the homology of the DNA sequences that encode these
30   antigens was examined. At conditions of moderate
stringency, no hybridization was observed between

-20-

the insert DNA and Y3147 (an M. tuberculosis 19kD
clone) and Y3179 (an M. leprae 18kD clone). This
indicates no significant homology between the DNA
sequences of the insert DNAs of these two clones.
05   This result suggests that the M. tuberculosis 19kD
and the M. leprae 18kD proteins are unlikely to be
homologous.

    As a result of the work described, recombinant
DNA clones encoding five major protein antigens of
10   M. tuberculosis were isolated through the use of an
extensive collection of well-characterized murine
monoclonal antibodies. These five proteins were
also found to be major antigens in the rabbit
humoral immune response to M. tuberculosis. One of
15   these antigens, the 65kD protein, is shared with
another other mycobacterial pathogen M. leprae.

    Several lines of evidence indicate that the
65kD antigen is among the most immunodominant of the
protein antigens of M. tuberculosis. Eleven of the
20   25 different M. tuberculosis and M. leprae mono-
clonal antibodies examined in this study recognized
the 65kD recombinant antigen from one or both
mycobacteria. In addition, almost half of the
recombinant DNA clones isolated with rabbit poly-
25   clonal anti-M. tuberculosis sera express the 65kD
antigen, reflecting the predominance of antibody to
this antigen in these sera.

    Considerable evidence indicates that the 65kD
antigen plays an important role in the human re-
30   sponse to tuberculosis. Antibodies directed against
this protein can be detected in the serum of

-21-

patients with tuberculosis. The 65kD antigen is
present in purified protein derivatives (PPD's) of
M. tuberculosis, M. bovis, and other mycobacteria.
Thole, J.E.R. et al., Infection Immunity, 50:800-806
05   (1985). Finally, helper T cell clones reactive with
recombinant 65kD antigen have been isolated from
patients with tuberculosis, indicating that this
antigen is involved in the cell-mediated as well as
the humoral immune response to tuberculosis.
10   Among the major antigens of the leprosy
bacillus, the 65kD antigen appears to elicit anti-
body and T cell responses similar to those observed
for the M. tuberculosis antigen. Both serum anti-
bodies and T cells directed against the 65kD M.
15   leprae antigen have been observed in patients with
leprosy. Britton, W.J. et al., Journal of
Immunology, 135:4171-4177 (1985); Mustafa, A.S. et
al., Nature, 319:63-66 (1986). In addition, T cell
clones from leprosy patients have been found to
20   respond to recombinant 65kD protein of M. bovis, as
well as to PPD's from both M. bovis BCG and M.
leprae. Emmrich, F. et al., Journal of Experimental
Medicine, 163:1024-1029 (1986); Shankar, P. et al.,
Journal of Immunology, 136:4255-4263 (1986). It is
25   interesting to note that in vaccine trials in Asia
and Africa, BCG provided significant protection
against leprosy, ranging from 20% to 80%. Fine, P.,
Tubercle, 65:137-153 (1984). An intriguing possi-
bility is that the M. bovis BCG 65kD antigen is
30   involved in engendering the immune protection

-22-

provided by this vaccine against M. leprae, as well
as against M. tuberculosis.

In addition to the 65kD antigen, there is
evidence that the 19kD and 71kD antigens of M.
05 tuberculosis may be particularly important in the
immune response to this bacillus. Helper T cell
clones from tuberculosis patients have been isolated
which respond to the recombinant 19kD protein. The
71kD antigen is recognized by the humoral immune
10 system of both mice and rabbits, and antibody to
this antigen has been shown to be a prominent
component of hyperimmune anti-M. tuberculosis rabbit
sera.

VIII.   Isolation of DNA Clones for Genes Encod-
15   ing Proteins of Additional Mycobacteria
Using the procedures described above for
isolation of genes encoding M. tuberculosis
proteins, genes encoding proteins of additional
mycobacteria were isolated. DNA clones containing
20 genes encoding the following proteins were isolated:

Mycobacterium	Protein	Clone
M. bovis BCG	71kD	PL1-101
	65kD	PL1-105
	19kD	PL1-501
	14kD	PL1-502
M. smegmatis	65kD	PL1-206
M. avium	65kD	PL1-401
M. africanum	65kD	PL1-301

25

-23-

For purposes of comparison, genes encoding the following proteins were isolated for <u>M. tuberculosis</u> and <u>M. leprae</u>:

Mycobacterium	Protein	Clone
<u>M. tuberculosis</u>	71kD	Y3272
	65kD	Y3150
	19kD	Y3147
	14kD	Y3248
<u>M. leprae</u>	65kD	

05

The following strains were used for this purpose:

Species	Isolate
<u>M. leprae</u>	Armadillo isolate (WHO)
<u>M. tuberculosis</u>	Erdmann strain
<u>M. africanum</u>	African clinical isolate
<u>M. bovis</u> BCG	Danish vaccine strain
<u>M. smegmatis</u>	MC^2-6
<u>M. avium</u>	AIDS patient isolate

10

15

DNA from recombinant lambda gt11 clones was isolated, as described above, and mapped with restriction endonucleases, using standard techniques. Davis, R.W. <u>et al.</u>, <u>Advanced Bacterial Genetics: A Manual for Genetic Engineering</u>, Cold Spring Harbor (1980).

20

Figure 4 presents a comparison of the restriction maps for four genes of <u>M. tuberculosis</u> with the restriction maps for four genes of <u>M. bovis</u> BCG which encode proteins of the same molecular weight. As is evident from the figure, in each case, the

25

-24-

restriction sites on the two genes (e.g., those on
the M. tuberculosis gene and those on the M. bovis
gene which encodes a protein of the same molecular
weight) are essentially identical. This indicates
05   that the sequence of the genes of the two myco-
bacteria (at least those encoding these four pro-
teins) are very similar and, therefore, the proteins
they encode are also very similar.

Figure 5 presents a comparison of the restric-
10   tion map for the gene encoding the 65kD protein for
the six mycobacteria. As is evident, the restric-
tion maps for the genes encoding the 65kD protein of
M. tuberculosis, M. africanum, M. bovis BCG, M.
smegmatis and M. avium are essentially identical.
15   The fact that there is no detectable difference
among these mycobacteria at the level of the
restriction map is an indication that, at least at
this level, the encoded proteins are the same.

As is also evident, the map of the M. leprae
20   65kD gene has several identical restriction sties in
common with those of the other mycobacteria; it also
has two sites not found in the other genes and lacks
three sites present in the others. This indicates
that, at the level of the restriction map, there are
25   similarities in the DNA (and the encoded protein).
In addition, however, there are differences apparent
at this level.

IX. Diagnostic, Therapeutic and Preventive
Applications
30   The isolation of genes encoding major protein
antigens of M. tuberculosis makes it possible to

-25-

address problems which presently exist in diagnosing
treating and preventing tuberculosis. Isolation of
genes encoding proteins of other mycobacteria, such
as M. bovis BCG, M. africanum, M. smegmatis and M.
05   avium makes it possible to address similar problems
in diseases which they cause.
.   The nucleotide sequence of three of the five
genes has been determined. The sequence of the
remaining genes can be determined using well-known
10   methods, such as that of Sanger et al. Sanger, F.
et.al., Proceedings of the National Academy of
Sciences, USA, 74:5463-5467 (1977). The amino acid
sequence of each of the immunodominant proteins has
been deduced from the nucleotide sequence of the
15   three genes and can be done for the others.
        Identification and characterization of the
genes for major tuberculosis protein antigens and of
the proteins themselves make it possible to develop
improved reagents for diagnosis and immuno-
20   prophylaxis of tuberculosis. Proteins antigens
encoded by an entire gene, or amino acid sequences
(e.g., peptides, protein fragments) which make up
the antigenic determinant of a M. tuberculosis
antigen (i.e., M. tuberculosis-specific antigenic
25   determinants) may be used in serodiagnostic tests
and skin tests. Such antigens would be highly
specific to the tuberculosis bacillus and the tests
in which they are used would also be highly
specific. Highly specific serological tests would
30   be of great value in screening populations for

-26-

individuals producing antibodies to M. tuberculosis-
specific antigenic·determinants; in monitoring the
development of active disease in individuals and in
assessing the efficacy of treatment. As a result,
05   early diagnosis of tuberculosis will be feasible,
thus making it possible to institute treatment at an
early stage of the disease and, in turn, to reduce
the likelihood it will be transmitted.

As a result of the work described, it is also
10   possible to determine which segment(s) of the M.
tuberculosis antigen is recognized by M.
tuberculosis-specific T cells. A mixture of pep-
tides recognized by helper T cells can serve as a
specific skin test antigen useful in assessing
15   immunological status (delayed hypersensitivity) of
infected individuals and those with whom they come
in contact. This specific skin test antigen would
be useful in evaluating rapidly the immunological
efficacy of anti-tuberculosis vaccines.

20   It is reasonable to expect that the products
encoded by M. tuberculosis genes, particularly those
shown to be recognized by helper T.cells, are
themselves immunogenic and thus useful components of
vaccines against tuberculosis. These products
25   include proteins and portions of such proteins
(e.g., polypeptides and peptides). For example, one
approach to vaccine development is the introduction
of genes encoding products (e.g., polypeptides)
which provide immunological protection into viruses
30   such as vaccinia virus, or bacteria, such as
cultivatable mycobacteria, thus producing a vaccine

-27-

capable of engendering long-lasting and very
specific immunity. The genes encoding five
immunodominant protein antigens of the tuberculosis
bacillus, described herein, are useful for that
05   purpose; genes encoding the 65kD, 19kD and 71kD
antigens, or a portion thereof, are particularly
valuable in vaccine construction.

    Because of the similarities in the DNA encoding
similarly-sized proteins and, thus, of the encoded
10   proteins themselves, it is possible that, for
example, a vaccine effective against two or more of
the mycobacteria can be produced.

EXEMPLIFICATION

Isolation and Analysis of Recombinants Expressing
15   the 65kD M. tuberculosis Antigen
    The recombinant DNA library of M. tuberculosis
genomic DNA fragments in the lambda gt11 vector was
constructed as described above. Recombinant phage
lambda RY3143 and lambda RY3146 were used. Young,
20   R.A. et al., Proceedings of the National Academy of
Sciences, USA, 82:2583-2587 (1985). Subclones of
the mycobacterial DNA inserts in these recombinant
phage were constructed in pUC19 or M13mp9 vectors
using standard recombinant DNA techniques. Messing,
25   J. and J. Viera, Gene, 19:269-276 (1982). Maniatis,
T. et al., Molecular Cloning: A Laboratory Manual,
Cold Spring Harbor Laboratory, Cold Spring Harbor,
NY (1982).

-28-

Monoclonal antibodies specific for the 65kD
antigen were obtained from the Immunology of
Tuberculosis Scientific Working Group under a grant
from the WHO/World Bank/UNDP Special Program for
05 Vaccine Development. These antibodies included
IT-13 (WTB-78), IT-31 (SA2D5H4), and IT-33 (MLIIH9).
Coates, A.R.M. et al., Lancet, 2:167-169 (1981).
Gillis, T.P. and T.M. Buchanon, Immunology,
37:172-178 (1982). Anti-B-galactosidase antibodies
10 were purchased from CooperBiomedical. Polyclonal
rabbit antisera directed against a sonicate of M.
tuberculosis strain H37Rv were elicited as described
by Minden and co-workers. Minden, P. et al.,
Infect. Immun., 46:519-525 (1984). Results are
15 shown in Table 2.

-29-

TABLE 2:   PATTERNS OF ANTIBODY REACTIVITIES[a]

Number of Clones	Reactivity with Antibodies		
	IT-13	IT-31	IT-33
27	+	+	+
1	+	+	+
2	+	-	+
3	-	+	+
1	+	-	-
2	-	+	-
2	-	-	+

[a]: Recombinant clones expressing antigens reactive
with the 65kD antigen specific monoclonal antibodies
IT-13, IT-31, and IT-33 were isolated as described
above. For the initial screen, a pool of the three
antibodies was used; it contained a 1:1000 dilution
of each antibody to screen a total of about 8×10^5
recombinant phage from the lambda gt11-M.
tuberculosis library. To determine which monoclonal
antibody reacted with which of the 38 plaque-
purified recombinants, about 100 pfu of each
recombinant phage were inoculated in small spots on
a lawn of Y1090. The phage were allowed to grow and
induced to synthesize the foreign proteins as
described previously. The filters were then reacted
with a 1:1000 dilution of one of the monoclonal
hybridoma antibodies as described above.

-30-

The lambda gt11-<u>M. tuberculosis</u> library was
screened with the monoclonal antibodies specific for
the 65kD antigen and clones reactive with them were
isolated essentially as described by Young <u>et al.</u>
05   Young, R.A. <u>et al.</u>, <u>Proceedings of the National</u>
<u>Academy of Sciences, USA</u>, <u>82</u>:2583-2587 (1985).
Briefly, for each 150mm LB plate, 0.6ml of a fresh
overnight culture of Y1090 was infected with 1-2
x10^5 plaque forming units of the library. After
10   3.5-4 hours of growth at 42°C, the plaques were
overlaid with a dry nitrocellulose filter which had
been saturated with 10mM isopropyl-B-D-thiogalacto-
pyranoside (IPTG). The plates were incubated an
additional 3.5-4 hours at 37°C and then removed to
15   room temperature and the position of the filters
marked. The filters were washed briefly in TBST (50
mM Tris-HCl, pH 8, 150mM NaCl, 0.05% Tween 20) and
then incubated in TBST + 20% fetal calf serum.
After 30 minutes at room temperature, the filters
20   were transferred to TBST plus antibody. For the
initial screen, the antibody mix contained a 1:1000
dilution of IT-13, IT-31, and IT-33. The filters
were incubated with the antibody solution overnight
at 4°C with gentle agitation, washed in TBST and
25   reacted with biotinylated goat anti-mouse immuno-
globulin, the Vectastain ABC reagent, and developer
as described by the manufacturer (Vector
Laboratories). After the color had developed the
filters were washed with several changes of water
30   and air dried. Phage corresponding to positive
signals were twice plaque purified. To determine

-31-

which monoclonal antibodies reacted with which of
the recombinant phage, about 100 pfu of a purified
phage stock were inoculated in a small spot on a
lawn of Y1090 bacteria on an LB plate. The phage
05   were allowed to grow and induced to synthesize the
foreign proteins as described above. The filters
were then reacted with a 1:1000 dilution of one of
the monoclonal antibodies. The subsequent steps
were the same as for the initial screen.
10   Western blot assays were carried out as
follows: Cells containing phage or plasmids in
which the expression of the foreign sequences was
under the control of the E. coli lac gene regulatory
sequences were induced to synthesize the foreign
15   proteins by incubating the cells in the presence of
2.5mM IPTG for 2 hours. Crude lysates of cells
expressing lambda gt11 recombinants were made as
described in Huynh et al. Huynh, T.V. et al., In:
DNA Cloning Techniques: A Practical Approach, (D.
20   Glover, ed.) IRL Press, Oxford, Vol. 1, pp. 49-78
(1985). Crude lysates of cells expressing plasmid
encoded proteins were made by harvesting cells from
overnight cultures and resuspending the cells in 10
mM Tris pH7.5/10 mM EDTA containing 100 ug
25   lysozyme/ml. After 10 minutes at room temperature,
SDS was added to a final concentration of 0.5%. A
protease inhibitor (Trasylol, Boehringer Mannheim)
was added to all crude lysates at a final concen-
tration of 0.3%. The crude protein preparations
30   were electrophoresed on 10% polyacrylamide-SDS
Laemmli gels and the separate proteins electrophor-

-32-

etically transferred to nitrocellulose. Laemmli,
U.K., Nature, 227:680-685 (1970). Towbin, H. et
al., Proceedings of the National Academy of
Sciences, USA, 76:4350-4354 (1979). The immobilized
05   proteins were reacted with a 1:1000 dilution of
monoclonal antibody IT-13 in TBST overnight at 4°C.
The nitrocellulose filters were then washed, reacted
with peroxidase-conjugated goat anti-mouse immuno-
globulin, and developed as described by Niman and
10   co-workers. Niman, H.L. et al., Proceedings of the
National Academy of Sciences, USA, 80:4949-4953
(1983).

The sequences of 5'-end-labeled restriction
fragments of the mycobacterial DNA were determined
15   by a modification of the partial chemical
degradation technique of Maxam and Gilbert. Brow,
M.A.D. et al., Mol. Biol. Evol., 2:1-12 (1985).
Maxam, A.M. and W. Gilbert, Proceedings of the
National Academy of Sciences, USA, 74:560-564
20   (1976). For the M13/dideoxy sequencing studies,
Sau3AI fragments from the mycobacterial DNA inserts
were subcloned into the BamHI site of M13mp9. Phage
DNA was isolated from the M13 recombinants and
subjected to the dideoxy chain termination
25   sequencing reactions. Biggin, M.D. et al.,
Proceedings of the National Academy of Sciences,
USA, 80:3963-3965 (1983). Sanger, F. et al.,
Journal of Molecular Biology, 143:161-178 (1980).
The products of the sequencing reactions were
30   electrophoresed on 6% acrylamide/7M urea/0.5-2.5 x
TBE gradient sequencing gels. The gels were dried

-33-

under vacuum and exposed to Kodak XRP-1 film. The
nucleotide sequences were-determined independently
for both strands of the mycobacterial DNA.

   Computer-aided analyses of the nucleic acid
05 sequences and deduced protein sequences were
performed using the Databases and programs provided
by the Nucleic Acid and Protein Identification
Resources of the National Institutes of Health as
well as the programs of Chou and Fasman and Hopp and
10 Woods. Chou, P.Y. and G.D. Fasman, Adv. Enzym.,
47:45-148 (1978). Hopp, T.P. and K.P. Woods,
Proceedings of the National Academy of Sciences,
USA, 78:3824-3828 (1981). The nucleotide sequence
of the region containing the M. tuberculosis 65kD
15 gene and the deduced amino acid sequence of the two
long open reading frames are represented in Figure
8.

   B-galactosidase assays were also carried out.
Cells were grown in LB broth or LB broth plus 2.5mM
20 IPTG to an OD_{600} of about 0.3. Crude lysates were
made and b-galactosidase activity assayed as
described by Miller. Miller, J.H., Experiments in
Molecular Genetics, Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY (1972).

25 Equivalents
   Those skilled in the art will recognize or be
able to ascertain, using no more than routine
experimentation, many equivalents to the specific
materials and components described herein. Such
30 equivalents are intended to be encompassed in the
scope of the following claims.

-34-

CLAIMS

1.   Isolated DNA encoding an immunogenic protein
      antigen of <u>Mycobacterium tuberculosis</u>.

2.   DNA of Claim 1 selected from the group consist-
05   ing of DNA encoding <u>Mycobacterium tuberculosis</u>
      protein antigens of molecular weight 71kD,
      65kD, 19kD, 14kD and 12kD.

3.   Isolated DNA encoding an immunodominant protein
      antigen of <u>Mycobacterium tuberculosis</u>, the
10   protein antigen having a molecular weight of
      approximately 65kD and recognized by a
      monoclonal antibody selected from the group
      consisting of: IT-31; C1.1; IIH9; IIC8; T2.3;
      Y1-2; SA2.D7C and IT-13.

15 4. Isolated DNA encoding an immunodominant protein
      antigen of <u>Mycobacterium tuberculosis</u>, the
      protein antigen having a molecular weight of
      approximately 19kD and recognized by a mono-
      clonal antibody selected from the group con-
20   sisting of: IT-10; IT-12; IT-16; and IT-19.

5.   Isolated DNA encoding an immunodominant protein
      antigen of <u>Mycobacterium tuberculosis</u>, the
      protein antigen having a molecular weight of
      approximately 71kD and recognized by the
25   monoclonal antibody IT-11.

-35-

6. Isolated DNA encoding an antigenic determinant
    of <u>Mycobacterium tuberculosis</u> protein.

7. DNA of Claim 6 which encodes an antigenic
    determinant selected from the group consisting
05 of antigenic determinants of <u>Mycobacterium</u>
    <u>tuberculosis</u> proteins of molecular weight 71kD,
    65kD, 19kD, 14kD and 12kD.

8. Isolated DNA encoding an amino acid sequence of
    an antigenic determinant of <u>Mycobacterium</u>
10 <u>tuberculosis</u> protein, said protein having a
    molecular weight of approximately 65kD.

9. Isolated <u>Mycobacterium tuberculosis</u> DNA encod-
    ing an immunodominant protein antigen having a
    molecular weight of approximately 65kD, said
15 DNA selected from the group consisting of:
    a.   the DNA insert of clone Y3141;
    b.   the DNA insert of clone Y3143;
    c.   the DNA insert of clone Y3150;
    d.   the DNA insert of clone Y3253; and
20 e.   the DNA insert of clone Y3262.

10. A protein antigen encoded by DNA of Claim 9.

11. A protein antigen of Claim 10, wherein the
    protein antigen is recognized by a monoclonal
    antibody selected from the group consisting of
25 IT-31; C1.1; IIH9; IIC8; T2.3; Y1-2; SA2.D7C
    and IT-13.

12. Isolated DNA having a nucleotide sequence selected from the group consisting of: a) the nucleotide sequence represented in Figure 6, or a portion thereof; b) the nucleotide sequence represented in Figure 7, or a portion thereof; and c) the nucleotide sequence represented in Figure 8, or a portion thereof.

13. A protein or a peptide selected from the group consisting of: a) proteins or peptides encoded by the nucleotide sequence represented in Figure 6, or a portion thereof; b) proteins or peptides encoded by the nucleotide sequence represented in Figure 7, or a portion thereof; and c) proteins or peptides encoded by the nucleotide sequence represented in Figure 8, or a portion thereof.

14. A peptide having the amino acid sequence of an antigenic determinant of Mycobacterium tuberculosis protein, said antigenic determinant being unique to Mycobacterium tuberculosis protein.

15. A peptide encoded by isolated Mycobacterium tuberculosis DNA, said peptide recognized by helper T cells.

16. A peptide encoded by the Mycobacterium tuberculosis DNA insert of clone Y3150 or a portion of said DNA insert.

17. Isolated DNA encoding a protein of <u>Myco-
     bacterium africanum</u> the protein having a
     molecular weight of 65kD.

18. Isolated DNA encoding a protein of <u>Myco-
     bacterium avium</u>, the protein having a molecular
     weight of 65kD.

19. A vaccine comprising DNA encoding <u>Mycobacterium
     tuberculosis</u> protein in a recombinant vaccine
     vector capable of expressing said DNA.

20. A vaccine of Claim 19 in which the recombinant
     vaccine vector is vaccinia virus or cultivat-
     able mycobacteria.

21. A vaccine of Claim 20 in which the DNA encodes
     the 65kD <u>Mycobacterium tuberculosis</u> protein
     recognized by the monoclonal antibody IT-13, or
     a portion of said protein.

22. A vaccine comprising DNA encoding an antigenic
     determinant unique to <u>Mycobacterium tubercu-
     losis</u> cultivatable mycobacteria capable of
     expressing said DNA.

23. A method of detecting antibody against <u>Myco-
     bacterium tuberculosis</u> in a biological fluid,
     comprising the steps of:
          a)   incubating an immunoadsorbent com-
     prising a solid phase to which is attached

-38-

immunodeterminant <u>Mycobacterium tuberculosis</u>
protein with a sample of the biological fluid
to be tested, under conditions which allow the
anti-<u>Mycobacterium tuberculosis</u> antibody in the
05   sample to bind to the immunoadsorbent;

b)   separating the immunoadsorbent from
the sample; and

c)   determining if antibody is bound to
the immunoadsorbent, as an indication of
10   anti-<u>Mycobacterium tuberculosis</u> in the sample.

24.   A method of Claim 23 in which the <u>Mycobacterium</u>
<u>tuberculosis</u> protein attached to the solid
phase has a molecular weight of approximately
65kD.

15   25.   A method of detecting antibody against <u>Myco-</u>
<u>bacterium tuberculosis</u> in a biological fluid,
comprising the steps of:

a)   incubating an immunoadsorbent com-
prising a solid phase to which is attached a
20   peptide having the amino acid sequence of an
antigenic determinant of <u>Mycobacterium</u>
<u>tuberculosis</u> protein with a sample of the
biological fluid to be tested, under conditions
which allow antibody against <u>Mycobacterium</u>
25   <u>tuberculosis</u> to bind to the immunoadsorbent;

b)   separating the immunoadsorbent; and

c)   determining if antibody is bound to
the immunoadsorbent, as an indication of the

-39-

presence of the antibody against <u>Mycobacterium</u> <u>tuberculosis</u> in the sample.

26. A method of Claim 25 in which the peptide has the amino acid sequence of an antigenic deter-
05   minant which is unique to <u>Mycobacterium</u> <u>tuberculosis</u> protein.

27. A kit useful in detecting antibody against <u>Mycobacterium tuberculosis</u> in a biological fluid, comprising a collection of reagents for
10   immunoassay of said antibody, said collection of reagents a solid phase to which is attached immunodeterminant <u>Mycobacterium tuberculosis</u> protein or a peptide having the amino acid sequence of an antigenic determinant of
15   <u>Mycobacterium tuberculosis</u>.

FIGURE 1

FIGURE 1 (Cont'd)

3/43

FIG.2

4/43
FIG.3
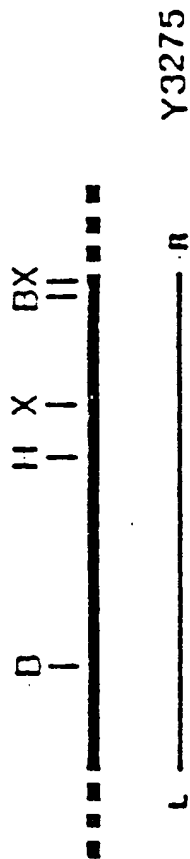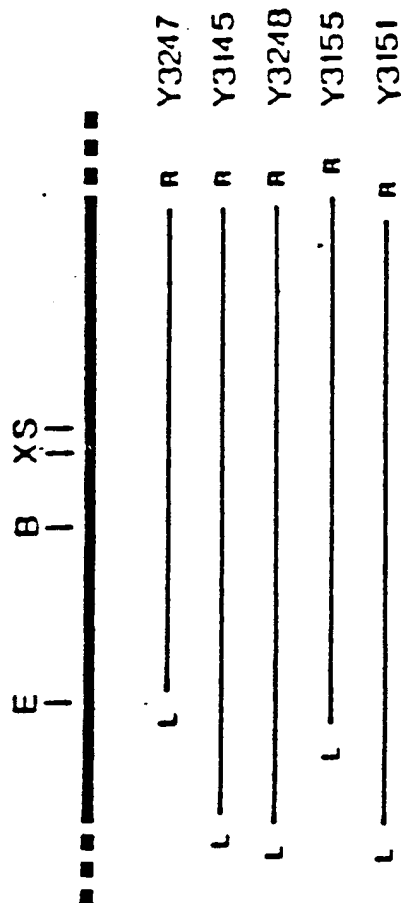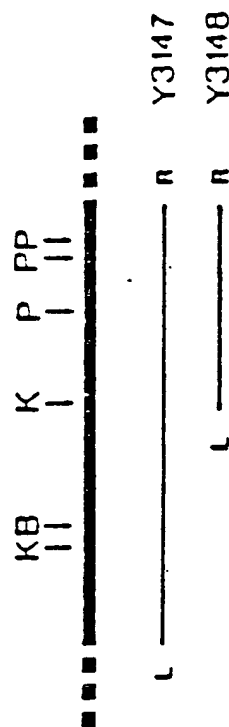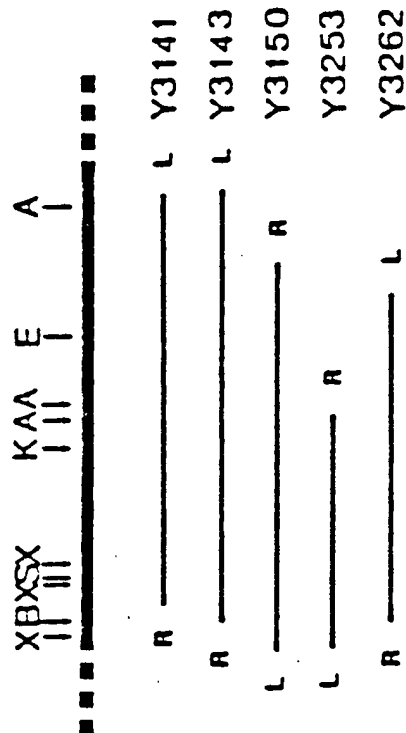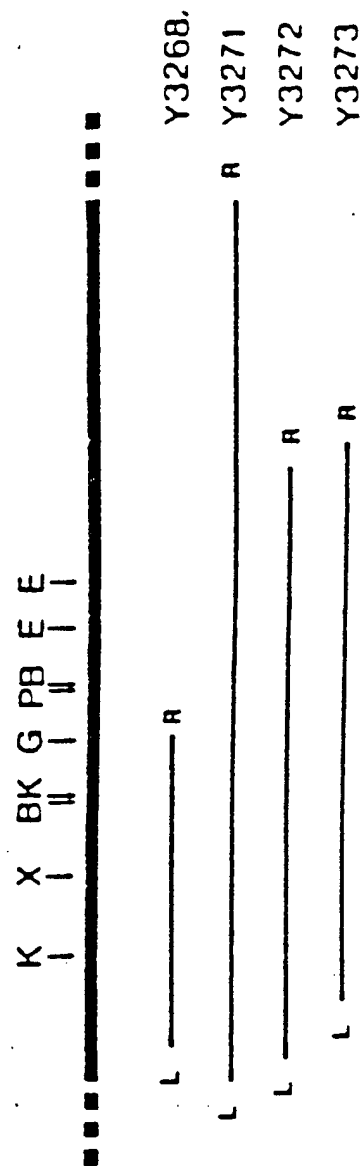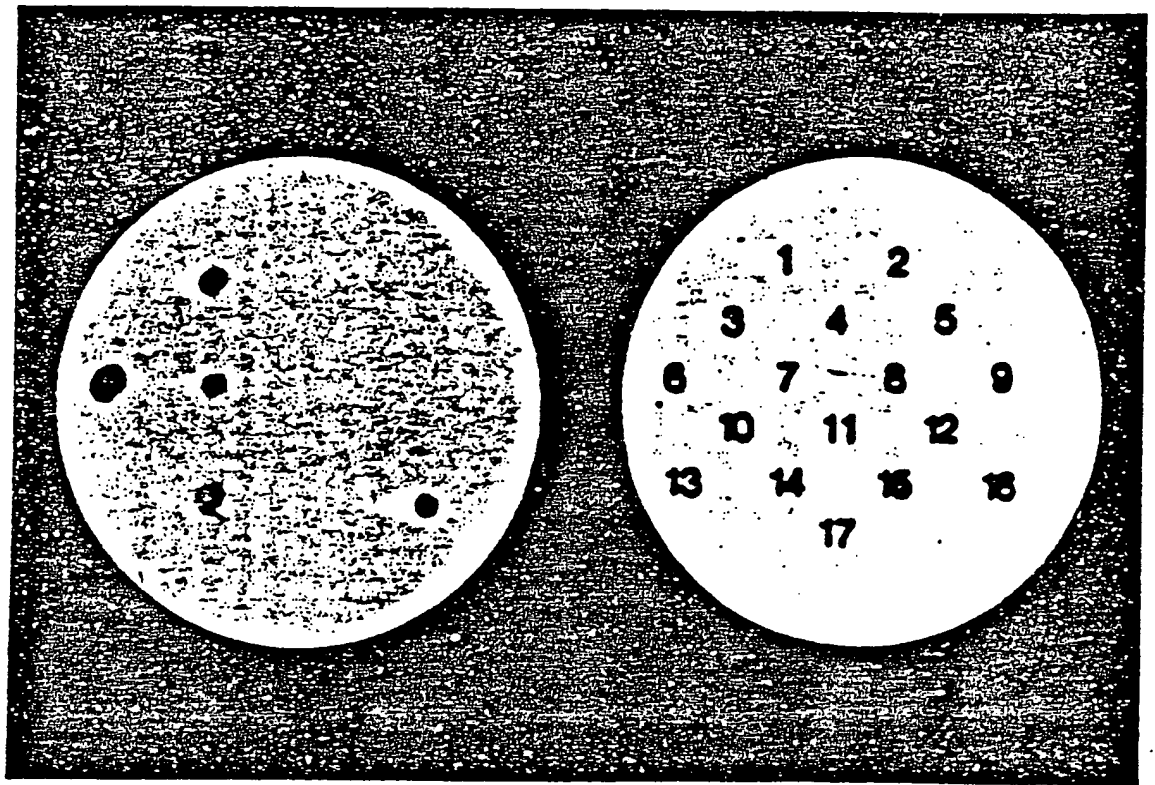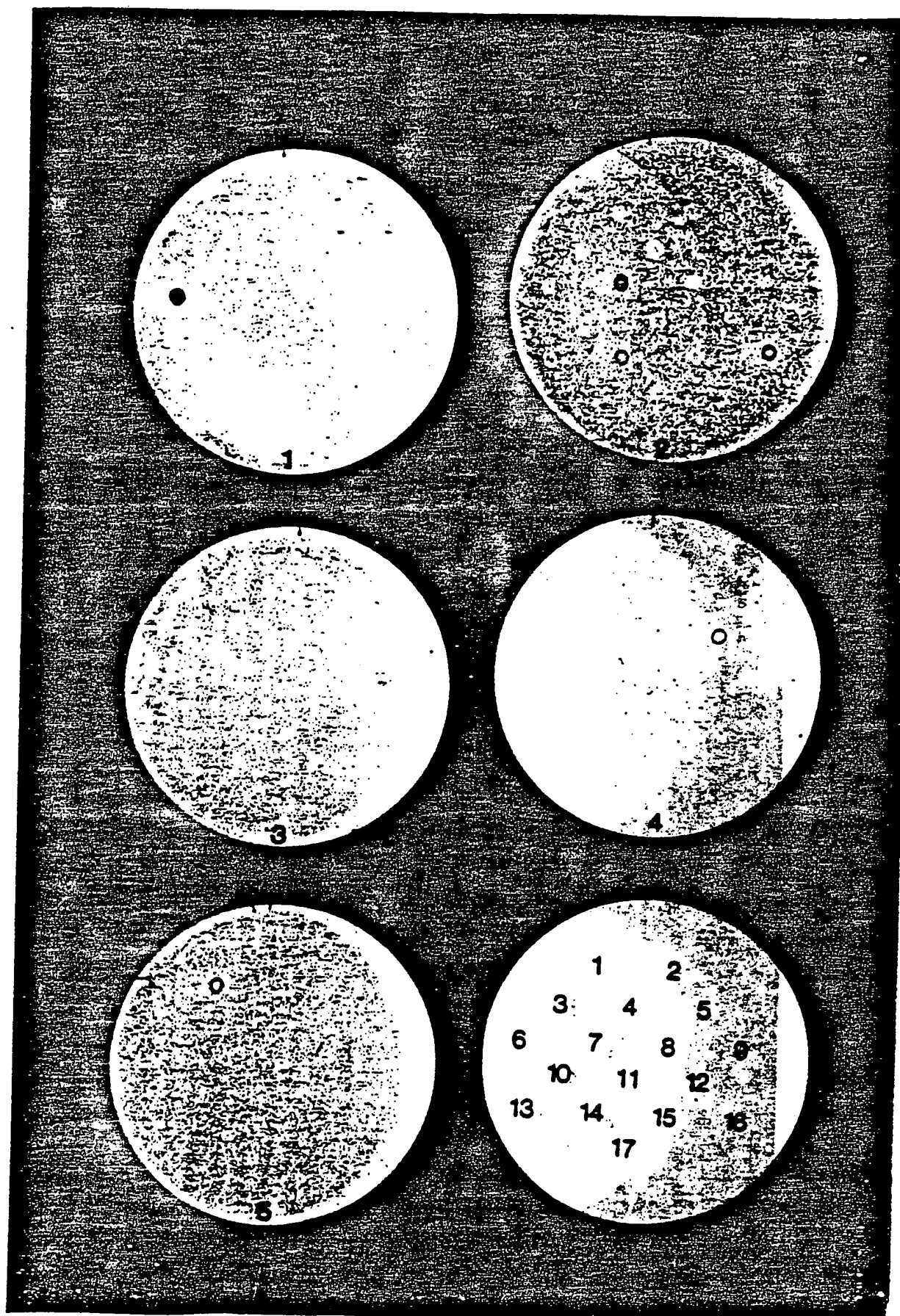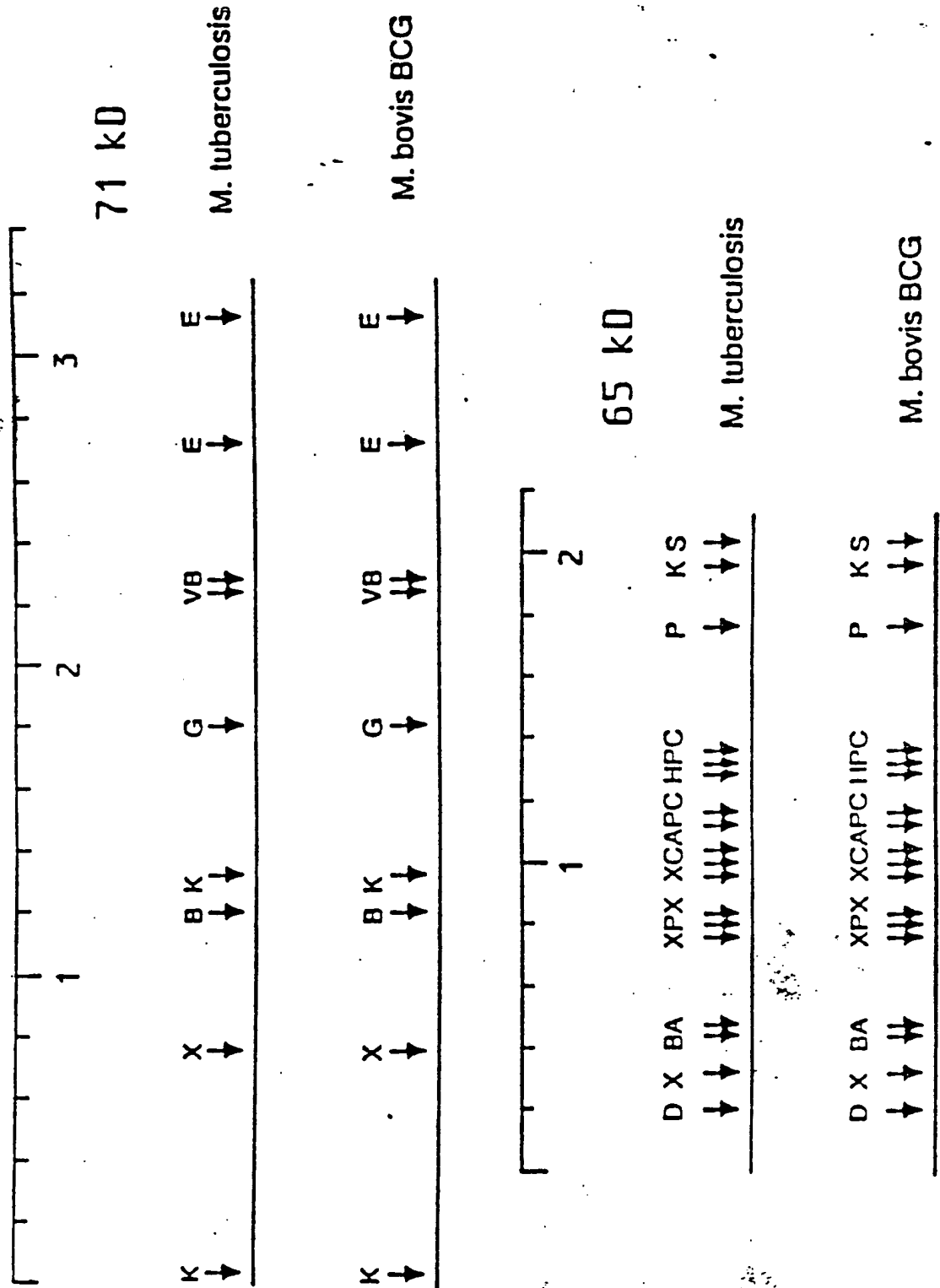
FIGURE 4

FIGURE 4 (Cont'd)

7/43
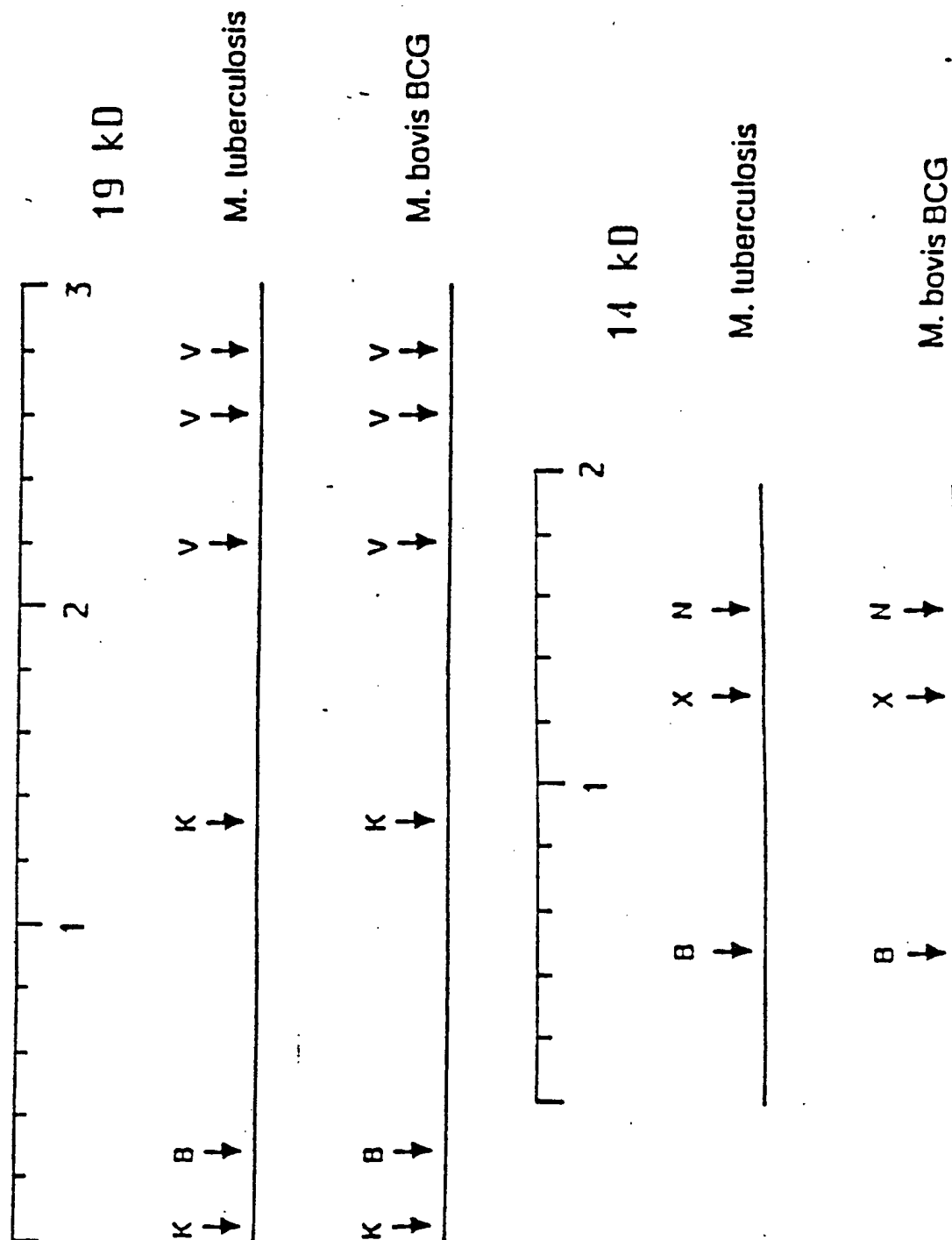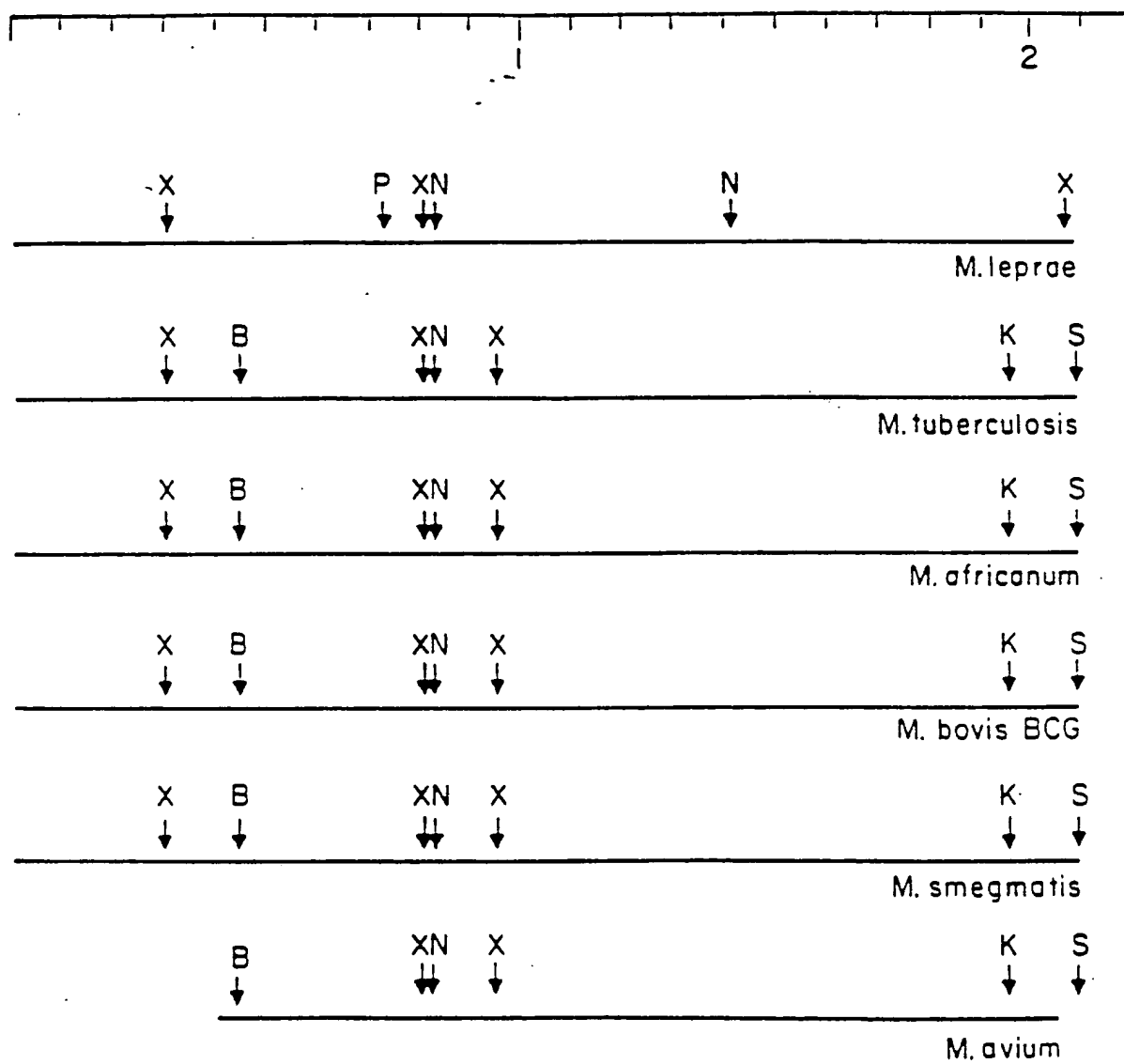
FIG. 5

FIGURE 6

F R V N R L G E I A R P G A R I A H Q G
S V S I A S V R * P D Q A R G S R T K A
P C Q S P R * D S P T R R A D R A P R R
TTCCGGTGTCAATCGCCTCGGTGTGAGATAGCCCGACCAGGCGGATCGCGCCACCAAGGC
          10   20   30   40   50   60
AAGGCACAGTTAGGCGGAGCCACTCTATCGGGTCCTGGTCCGCGCCTAGCGCGTGGTTCCG
E T D I A E T L Y G S W A R P D R V L A
G H * D G R H S L G V L R A S R A G; L R
R T L R R P S I A R G P A R I A C W P A

A A Q G Q Q V V Q V M R G V F G H A Q R
P R R V S R L C R L C V A F S A M H N A
R A G S A G C A G Y A W R F R P C T T R
GCCGCGCAGGGTCAGCAGGTTGTGCAGGTTATGCGTGGCGTTTTCGGCCATGCACAACGC
          70   80   90   100   110   120
CGGCGCGTCCCAGTCGTCGTCCAACACGTCCAATACGCCAAAAGCCGGTACGTGTTGCG
G R L T L N H L N H T A N E A M C L A.
R A P D A P Q A P * A H R K R G H V V R
A C P * C T T C T I R P T K P W A C R A

FIGURE 6 (CONT'D)

```
A E A R E I E V I L R R G L G A R G H L
R K R G K * K C I S A G A S V P G A I W
G S A G N R S A S P Q G P R C P G P S G
GCGGAAGGCGGGGAAATAGAAGTGCATCTCCGCAGGGCCTCGGTGCCCGGGCCATCTG
          130       140       150       160       170       180
CGCCTTCCGCCCTTTATCTTCACGTAGAGGCGTCCCGGAGCCACGGGCCCGGTAGAC
R F R F Y F H M E A P A E T G P A M Q
P L A P E L L A D G C P G R H G P G D P
S A R S I S T C R R L P R P A R P W R S


E L D L H P V D G V G L P G L G D V D R
N S I S T P S M V W V S P V S V M S T V
T R S P R R W C G S P R S R * C R P S
GAACTCGATCTCCACCCCGTCGATGGTGGGTCTCCCGGTCTCGGTGATGTCGACCGT
          190       200       210       220       230       240
CTTGAGCTAGAGGTGGGGCAGCTACCACCCCAGAGGGCCAGAGCCACTACAGCTGGCA
F E I E V G D I T H T E G T E T I D V T
V R D G R R H H P D G R D R H H R G D
S S R W G T S P T P R G P R P S T S R R
```

FIGURE 6 (CONT'D)

```
R H D E R N L T G R Q C L P E A A A D V
G T T S E T S P V D S V C P R Q P T C
A R R A K P H R S T V S A R G R S R R A
CGGCACGACGAGCGGAAACCTCACCGGTCGACAGTGTCTGCCCGAGGCCGCAGCCGACGTG
    250   260   270   280   290   300
GCCGTGCTGCTCGCTTTGGAGTGGCCAGCTGTCACAGACGGGCTCCGGCTCGGCTGCCAC
P V V L S V E G T S L T Q G L G C G V H
A R R A F G * R D V T D A R P R L R, R A
C S S R F R V P R C H R G S A A A S T G

P P E T A R Q H G A V H V A R T A H H R
P R R P R A N T V P Y M * P A R R I I A
P G D R A P T R C R T C S P H G A S S P
CCCCCGGAGACCGCGCCAACACGGTGCCGTACATGTAGCCCGCGGCCATCATCGC
    310   320   330   340   350   360
GGGGGCCTCTGGCGCCGGTTGTGCCACGGCATGTACATCGGGCGGCGTAGTAGCG
G R L G R A L V T G Y M Y G A R R M M A,
G P S R A G V R H R V H L G C P A D D G
G S V A R W C P A T C T A R V A C * R R
```

FIGURE 6 (CONT'D)

```
R A G V D V F L H G V R G E P L R R Q H
E P A * M F S C T A C A V N P S G A S T
  S R R C F P A R R A R * T P P A P A P
CGAGCCGGCGTAGATGTTTCCTGCACGGGCGTGCGTGGGGAGCCCTCCGGCGCCAGCAC
        370        380        390        400        410        420

GCTCGGCCGCATCTACAAAAGGACGTGCCCGCCACTTGGGAGGCCGCGGCGGTCGTG
S G A Y I N E Q V A H A T F G E P A L V
 L R R L H K G A R R A R H V G G A G A G
  A P T S T K R C P T R P S G R R R W C R


R H L S R V H V G L G G D A E H P T E M
A T F P A S T S A W V V T P S T P P K *
 P P F P R P R P G W * R R A P H R N D
CGCCACCTTCCCGTCCGTCCACGTGGGTGGTGACGCGCCGAGCACCCACCGAAATG
        430        440        450        460        470        480

GCGGTGGAAAGGGCGCAGGTGCAGCCCGACCCACCACTGCGGCTCGTGGGTGGCTTTAC
A V K G A D V D A Q T T V G L V G G F H
 G G K G R G R R G P H H R R A G W R E S
  W R E R T W T P R P P S A S C G V S I I
```

FIGURE 6 (CONT'D)

```
I D M A V G V D D R D H G A V G S A V G
  S T W L W V * M T A T T G R S A P R W A
  R H G C G C R * P P R G G R L R G G R
ATCGACATGGCTGTGGGTGTAGATGACCGCGACCACGGGGCGGTCGGCTCCGGCGGTGGGC
         490       500       510       520       530       540
TAGCTGTACCGACACCCACATCTACTGGCGCTGGTCCCGCCAGCCGAGGCCGCCACCCG
  D V H S H T Y I V A V V P R D A G R H A
  R C P Q P H L H G R P P R S R P P R P R
  S M A T P T S S R S W P A T P E A T P A


A I Q V Q R G G G H L G G H Q R V D D D
  R Y K S S A A A A T S V D T N G S M T I
  D T S P A R R P P R W T P T G R * R S
GCGATACAAGTCCAGCGCGGCGGGCACCTCGGTGACACCAACGGGTCGATGACGAT
         550       560       570       580       590       600
CGCTATGTTCAGGTCGCGCCGCCCGTGGAGCCACTGTGGTTGCCCAGCTACTGCTA
  R Y L D L A A A A V E T S V L P D I V I
  S V L G A R R G G R H V G V P R H R D
  I C T W R P P P W R P P C W R T S S S *
```

FIGURE 6 (CONT'D)

```
Q  P  S  V  T  L  N  E  A  D  I  G  D  I  E  S  A  D  L  I
  S  P  V  S  P  S  T  K  L  I  L  E  I  S  N  P  R  T  *  *
 A  Q  C  H  P  Q  R  S  *  Y  W  R  Y  R  I  R  G  P  D  R
CAGCCCAGTGTCACCCTCAACGAAGCTGATATTGGAGATATCGAATCCGGGACCTGATA
          610       620       630       640       650       660

GTCGGGTCACAGTGGGGAGTTGCTTCGACTATAACCTCTATAGCTTAGGCCTGGACTAT
 L  G  T  D  G  E  V  F  S  I  N  S  I  D  F  G  R  V  Q  Y
A  W  H  *  G  *  R  L  Q  Y  Q  L  Y  R  I  R  P  G  S  L
 G  L  T  V  R  L  S  A  S  I  P  S  I  S  D  A  S  R  I  S

D  A  R  H  H  L  V  E  A  L  F  R  G  Q  L  G  L  P  P  Q
 M  P  G  T  W  *  R  P  C  E  A  V  S  W  D  C  R  H  R
  C  P  A  P  P  G  R  G  P  V  S  R  S  A  G  I  A  A  T  G
GATGCCCGGCACCACCTGGTAGAGGCCCTGTTTCGGGGTCAGCTGGGATTGCCGCCACAG
          670       680       690       700       710       720

CTACGGGCCGTGGTGGTGGACCATCTCCCGGACAAAGCGCCAGTCGACCCTAACGGCGGTGTC
 I  G  P  V  V  Q  Y  L  G  Q  K  A  T  L  Q  S  Q  R  W  L
H  G  A  G  G  P  L  P  G  T  E  R  D  A  P  I  A  A  V  P
```

14/43

FIGURE 6 (CONT'D)

```
A R C W R T S A R N R P * S P N G G C A
  A G M H R C R G T V E K R V R V V V P
    L G C T D V G A A P S R N E Y A S L S H
      W D A P M S A R H R R E S T R C P T
GCTGGGATGCACCGATGTCGGCGCGCCGTCGAGAAACGAGTACGCGTCGTTGTCCCA
        730       740       750       760       770       780
CGACCCTACGTGGCTACAGCCGGCGCGGCCGTGGCAGCTCTTGCTCATGCGCAGCAACAGGGT
  S P H V S T P A A G D L F S Y A D N D W
    Q S A G I D A R C R R S V L V R R Q G V
      P I C R H R R P V T S F R T R T T G C

  H H A T I G S L D H T R G Q R G N E S A
    T T R P S A L I T H G D S A A M N P R
      P R D H R Q P * S H T G T A R Q * I R D
CACCACGCGACCATCGGCAGCCTTGATCACACACGGGGACAGCGGGGCAATGAATCCGCG
        790       800       810       820       830       840
GTGGTGCGCTGGTAGCCGTCGGAACTAGTGTGTGCCCCTGTCGCGCCGTTACTTAGGCGC
  V V R G D A K I V C P S L A I F G R
    G R S W R C G Q D C V P V A R C H I R S
      W A V M P L R S * V R P C R P L S D A I
```

FIGURE 6 (CONT'D)

```
 I  G  V  V  E  I  R  C  V  M  Q  R  *  R  V  F  T  V  C  R
 S  A  S  S  K  S  V  V  S  C  N  G  N  E  C  S  P  C  A  A
   R  R  R  N  P  L  C  H  A  T  V  T  S  V  H  R  V  P  P
ATCGGCGTCGTCGAAATCCGTTGTGTCATGCAACGGTAACGGAGTGTTCACCGTGTGCCGC
       850       860       870       880       890       900
TAGCCGCAGCAGCTTTAGGCAACACAGTACGTTGCCATTGCTCACAAGTGGCACACGGGCG
 D  A  D  D  F  D  T  D  H  L  P  L  S  H  E  G  H  A  A
 R  R  R  R  F  E  G  N  H  *  A  V  T  V  L  T  *  R  T  G  G
 P  T  T  S  I  R  Q  T  M  C  R  Y  R  T  N  V  T  H  R  R


 L  D  D  G  S  G  R  E  F  V  E  H  R  H  Y  I  A  T  T  T  V
 W  M  T  A  V  G  G  L  C  S  I  G  T  T  L  P  L  L  R  C
 G  *  R  Q  W  E  V  C  V  P  S  A  L  H  C  H  Y  Y  G  A
CTGGATGACGGCAGTGGGAGGTTTGTGTTCCATCGGCACTACTACGGTG
       910       920       930       940       950       960
GACCTACTGCCGTCACCCTCCAAACACAAGGTAGCCGTGATGATGCCAC
 Q  I  V  A  T  P  P  K  H  E  M  P  V  V  N  G  S  S  R  H
 P  H  R  C  H  S  T  Q  T  G  D  A  S  C  Q  W  *  P  A
 S  S  P  L  P  L  N  T  N  W  R  C  *  M  A  V  V  V  T  C
```

FIGURE 6 (CONT'D)

```
H   A   G   R   C   R   W   R   T   T   L   P   T   R   K   R   E   E   S   A
  T   P   V   D   A   V   G   E   P   R   Y   R   P   E   R   E   N   E   P   P
    R   R   *   M   P   L   A   N   H   A   T   D   Q   K   E   R   I   E   R   R
CACGCCGGTAGATGCCGTTGGCGAACCACGCCTACCGACCAGAAAGAGAGAATTTCCGCC
          970       980       990      1000      1010      1020

GTGGCGGCCATCTACGGCCAACCGCCTTGGTGCCGATGGCTGGTCTTTCTCTTAAAAGGCGG
V   G   T   S   A   T   P   S   G   R   *   R   G   S   L   S   F   K   G   G
  R   R   Y   I   G   N   A   E   W   A   V   S   W   F   S   L   I   K   R   R
    A   P   L   H   R   Q   R   V   V   S   G   V   L   F   L   S   N   E   A   A

A   P   P   R   A   L   L   T   R   I   L   P   K   R   S   S   M   P   M
  H   L   D   L   G   P   C   *   R   A   Y   C   R   S   G   P   Q   C   R   W
    T   *   T   S   G   P   A   N   A   H   T   A   E   A   V   L   N   A   D   G
GCACCTAGACCCTCGGGCCCTGCTAACGGCCATACTGCCGAAGCGGTCCTCAATGCCGATG
         1030      1040      1050      1060      1070      1080

CGTGGGATCTCGAGCCCGGACGATTGCCGTATGACGGCTTCGCCAGGAGTTACGGCTAC
C   R   S   R   P   G   Q   *   R   A   Y   Q   R   L   P   G   *   H   R   H   H
```

FIGURE 6 (CONT'D)

```
V * V E P G A L A C V A S A T R L A S P
G L G R A R S V R M S G F R D E I G I S

D R Y D R Q R S T G * S V D * R S R * P
T A T T G K G A Q G E A W T D G R G S R
P L R Q A K E H R [V K R G L T V A V A G
GACCGGCTACGACAGGCAAAGGAGCACAGGGTGAAGCGGGTCGGCGGTAGCCG
      1090      1100      1110      1120      1130      1140
CTGGCCGATGCTGTCCGTTCCTCGTGTCCCACTTCGCCACCTGACTGCCAGCGCCATCGGC
V A V V P L P A C P S A H V S P R P L R
G S R C A F S C L T F R P S V T A T A P
R * S L C L L V P H L T S Q R D R Y G S

E P P F W S Q V F P D V Q A T S R L Q E
S R H S G R S F R M F K Q Q V D Y R K
A A I L V A G L S G C S S N K S T T G S
GAGCCGCCATTCTGGTCGCAGGTCTTTCCGGATGTTCAAGCAACAAGTCGACTACAGGAA
      1150      1160      1170      1180      1190      1200
CTCGGCGGTAAGACCAGCGTCCAGAAAGGCCTACAAGTTCGTTGTTCAGCTGATGTCCTT
L R W E P R L D K R I N L C C T S * L F
A A M R T A P R E P H E L L L D V V P L
G G N Q D C T K G S T * A V L R S C S A
```

FIGURE 6 (CONT'D)

```
A V R P R P R Q A R R Q A P A P P P G R
R * D H D R G R H D G K P R R R L R A E
G E T T A A G T T A S P G A A S G P K
GCGGTGAGACCACGACCGGCAGGCACGGCAAGCCCCGGCGCCTCCGGGCCGA
         1210        1220       1230        1240       1250       1260
CGCCACTCTGGTGCTGGCGCCGTCCGTGCTGCCGGTTCGGGGCCGGGAGGCCCGGCT
R H S W S R P L C S P L G R R R R R' A S
P S V V V A A P V V A L G P A A E P G E
T L G R G R C A R R C A G A G G G G P R L

R S S S T V R T R T S P A P W C A Q P R
G R H R R' * G P E R H R L R G V H N R G
V V I D G K D Q N V T G S V V C T T A A
AGGTCGTCATCGACGGTAAGGACCAGAACGTCACCGGCTCCGTGGTGCACAACCGCGG
         1270        1280       1290        1300       1310       1320
TCCAGCAGTAGCTGCCATTCCTGGTCTTGCAGTGGCCGAGGCACCACGTGTTGGCGCC
P R' * R R Y P G S R' R S R P T C' L R P
T T M S P L S W E T V P E T H V V A A
D D D V T L V L V D G A G H H A C G R G
```

FIGURE 6 (CONT'D)

```
P A M S T S R S A G R R P A L P P C S P
  R Q C Q H R D R R G G D R H C R R A H R
    G N V N I A I G G A A T G I A A V L T D
CCGGCAATGTCAACATCGGATCGGGGGGCGGACCGGCATTGCCGCCGTGCTCACCG
      1330    1340    1350    1360    1370    1380
GGCCGTTACAGTTGTAGCCGCTAGCCGCCCCGCCCTGGCCGTAACGGCCACGAGTGGC
  R C H * C R S R P P S R C Q R R A * R
    P L T L M A I P P A A V P M A A T S V S
      A I D V D R D A P R R G A N G G H E G V

T A T L R R * S P L G S V T S T A S R W
  R Q P S G G E V R W A R * R Q R H A G
    G N P P E V K S V G L G N V N G V T L G
ACGGCAACCCTCCGGAGGTGAAGTCCGTTGGGCTCAACGGTCAACGGTCACGCGCTGG
      1390    1400    1410    1420    1430    1440
TGCCGTTGGGGAGGCCTTCCACTTCAGGCCATTGCCGCCAGTTGCCGCCAGTGCCGGACC
  R C G E P P S T F D T P S P L T L P T V S P
    P L G G S T F D T P S P L T L P T V S P
      A V R R L H L G N P E T V D V A D R Q S
```

FIGURE 6 (CONT'D)

```
D  T  R  R  A  P  D  R  V  T  P  R  Q  P  R  T  A  A  T  T
 I  H  V  G  H  R  T  G  *  R  L  G  N  Q  G  R  Q  P  L  Q
  Y  T  S  G  T  G  Q  G  N  A  S  A  T  K  D  G  S  H  Y  K
GATACACGTCGGGCCACCGGACAGGGTAACGCCTCGGCAACCAAGGACGGCAGCCACTACA
       1450      1460      1470      1480      1490      1500
CTATGTGCAGCCCGTGGCCCGTGTCCCATTGCGGAGCCCGTTGGTTCCTGCCGTCGGTGATGT
 I  C  T  P  C  R  V  P  Y  R  R  P  L  W  P  R  C  G  S  C
  Y  V  D  P  V  P  C  P  L  A  E  A  V  L  S  P  L  W  *  L
   V  R  R  A  G  S  L  T  V  G  R  C  G  L  V  A  A  V  V  L

 R  S  L  G  P  L  P  G  S  T  W  P  T  R  C  H  R  *  T  S
  D  H  W  D  R  Y  R  G  R  H  G  Q  P  D  V  T  G  E  Q  V
   I  T  G  T  A  T  G  V  D  M  A  N  P  M  S  P  V  N  K  S
AGATCACTGGGACCGCTACCGGGGTCGACATGGCCAACCCGATGTCAACCGGTGAACAAGT
       1510      1520      1530      1540      1550      1560
TCTAGTGACCCTGGCGATGGCCCCAGTGTACCGGTTGGGCTACAGTGGCCACTTGTTCA
 S  *  Q  S  R  *  R  P  R  C  P  W  G  S  T  V  P  S  C  T
  I  V  P  V  A  V  P  T  S  M  A  L  G  I  D  G  T  F  L  D
   D  S  P  G  S  G  P  D  V  H  G  V  R  H  *  R  H  V  L  R
```

FIGURE 6 (CONT'D)

```
R S K S R * P V P N L K R V D A G C E Q
V R N R G D L F L T * S V S M R A V N S
F E I E V T C S * P K A C R C G L * T A
CGTTCGAAATCGAGGTGACCTGTTCCTAACCTAAAGCCTGTGTCGATGCGGGCTGTGAACAG
      1570   1580   1590   1600   1610   1620
GCAAGCTTTAGCTCCACTGGACAAGGATTGGATTCGCACAGCTACGCCCGACACTTGTC
T R F R P S R N R V * L T D I R A T F L
N S I S T V Q E * G L A H R H P S H V A
E F D L H G T G L R E R T S A P Q S C R

R V G A G Q S G L A R R F E E R L P S V
A S E P G S Q A * R G D D S G C H P S
R R S R A V R P S A A T I R A V A I R Q
CGCGTCGGAGCCGGGCAGTCAGGCCTAGCGGCGATTCGAGCGGTTGCCATCCGTC
      1630   1640   1650   1660   1670   1680
GCGCAGCCTCGGCCCGTCCGGATCCGCCTGCTAAGCTCGCCAACGGTAGGCAG
A D S G P L * A * R P S S E L P Q W G D
R R L R A T L G L A A V I R A T A M R *
T P A P C D P R A R R R N S R N G D T L
```

FIGURE 6 (CONT'D)

```
K W Q P H R K L G I S G * A T H G D R S
S G N R T A N S V Y P G E L L T V I V P
V A T A P Q T R Y I R V S Y S R * S E F
AAGTGGCAACCGCACCCGCAAACTCGGTATATCCGGTGAGCTACTCACGGTGATCGTTCC
      1690    1700    1710    1720    1730    1740
TTCACCGTTGGCGTGGGCCGTTTGAGCCATATAGGCCCACTCGATGAGTGCCACTAGCAAGG
L P L R V A F E T Y G P S S S V T I T G
T A V A G C V R Y I R T L * E R H D N R
H C G C R L S P I D P H A V * P S R E T

V V R L D H S G D D R Q A E P G A T G L
L C A L T A E T I A R P S P V L P A W
C A P * P Q R R R S P G R A R C Y R L G
GTTGTGGCCTTGACCACAGCGGAGACGATCGCCAGGCCCGGTGCTACCGGCTTG
      1750    1760    1770    1780    1790    1800
CAACACGGGAACTGGTGTCGCCTCTGCTAGCCGGTCCGGGCCACGATGGCCGAAC
N H A K V V A S V I A L G L G T S G A Q
Q A G Q G C R L R D G P R A R II * R S P
T R R S W L P S S R W A S G P A V P K A
```

FIGURE 6 (CONT'D)

```
A  G  P  *  R  I  A  A  G  E  P  L  E  N  L  G  L  Q  R  G
   R  D  R  D  V  S  P  R  A  N  R  S  K  T  S  D  C  S  A  A
      G  T  V  Y  R  G  R  T  A  R  K  P  R  T  A  A  R  P
GCGGGACCGGTGACGTATCGCGGGCGAACCGCTCGAAAACCTCGGACTGCAGCGCGGC
      1810      1820      1830      1840      1850      1860
CGCCCTGGCACTGCATAGCGGCGCCCGCTTGGCGAGCTTTTGGAGCCTGACGTCGCGCCG
R  S  R  S  T  D  G  R  A  F  R  E  F  V  E  S  Q  L  A  A
   P  V  T  V  Y  R  R  P  R  V  A  R  E  G  R  V  A  R  G
      P  G  H  R  I  A  A  P  S  G  S  S  F  R  P  S  C  R  P  R

R  N  T  R  P  I  V  D  H  L  Q  H  D  V  R  R  P  G  A  Q
   G  I  P  G  P  L  S  I  T  C  S  T  T  C  V  G  P  V  L  K
      E  Y  P  A  H  C  R  S  P  A  A  R  R  A  S  A  R  C  S  S
CGGAATACCCGGCCCATTGTCGATCACCTGCAGCACGACGTGCGTCGGCCGGTGCTCAA
      1870      1880      1890      1900      1910      1920
GCCTTATGGGCCGGTAACAGCTAGTGGACGTCGTGCTGCACGCAGCCGGCCACGAGTT
P  I  G  P  G  N  D  I  V  Q  L  V  V  I  T  P  G  T  S  L
   S  Y  G  A  W  Q  R  D  G  A  A  R  R  A  D  A  R  H  E  L
      F  V  R  G  M  T  S  *  R  C  C  S  T  R  R  G  P  A  *  A
```

FIGURE 6 (CONT'D)

```
A  R  R  H  D  R  A  G  T  G  V  H  A  G  V  G  Q  Q  V  G
R  V  V  T  I  V  P  G  P  V  C  T  R  A  L  A  S  R  L  V
A  S  S  R  S  C  R  D  R  C  A  R  G  R  W  P  A  G  W  S
GCGGTCGTCACGATCGTGCCGGACCGGTGTGCACGCGGGCGTTGGCCAGCAGGTTGGT
      1930      1940      1950      1960      1970      1980
CGGCGCAGCAGTGCCTAGCACGGCCCTGCCCGCCCACACCGGTCGTCCAACCA
R  T  T  V  I  T  G  P  T  H  V  R  A  N  A  L  L  N  T
A  D  D  R  D  H  R  S  R  H  A  R  P  R  Q  G  A  P  Q  D
R  R  *  S  R  A  P  V  P  T  C  A  P  T  P  W  C  T  P  *

H  H  L  V  Q  P  C  R  I  T  R  D  D  H  R  E  G  G  Q  V
T  T  W  C  N  R  A  A  S  P  G  M  T  T  G  S  G  G  R  S
P  P  G  A  T  V  P  H  H  P  G  *  P  P  V  R  G  A  G  R
CACCACCTGGTGCAACCGTGCCGATCACCCGGATGACCACCGGTTCGGGGGGCAGGTC
      1990      2000      2010      2020      2030      2040
GTGGTGGACCACGTTGGCACGGGCGTAGTGGGCCCTACTGGTGCCCAAGCCCCGTCCAG
V  V  Q  H  L  R  A  A  D  G  P  I  V  V  P  E  P  P  L  D
G  G  P  A  V  T  G  C  *  G  P  H  G  G  T  R  P  A  P  R
W  R  T  C  G  H  R  M  V  R  S  S  W  R  N  P  P  C  T  S
```

FIGURE 6 (CONT'D)

```
E  R  P  L  V  I  W  S  G  N  M  S  V  A  D  R  V  N  R  K
 S  A  H  W  *  S  G  P  A  T  *  A  S  L  T  A  S  T  A  S
  A  P  T  G  D  L  V  R  Q  H  E  R  R  *  P  R  Q  P  Q  A
GAGCGGCCCACTGGTGATCTGGTCCGGCAACATGAGCGTCGCTGACCGGTCAACCGCAAG
    2050      2060      2070      2080      2090      2100
CTCGCGGGTGACCACTAGACCAGGCCGTTGTACTCGCAGCGACTGGCGCAGTTGGCGGTTC
L  A  W  Q  H  D  P  G  A  V  H  A  D  S  V  A  D  V  A  L
 A  G  V  P  S  R  T  C  C  S  R  R  Q  G  R  *  G  C  A
  R  G  S  T  I  Q  D  P  L  M  L  T  A  S  R  T  L  R  L  G

P  R  H  V  H  R  S  A  F  Q  R  P  P  P  R  V  Q  T  R  Q  Q
 R  D  M  S  T  G  P  R  S  S  G  R  P  E  S  R  R  A  S  S
  A  T  C  P  P  V  R  V  P  A  A  A  P  S  P  D  A  P  A  A
CCGGCGACATGTCCACCGGTCCGGCTTCCAGCGGCCGCCCGAGTCCCAGACGCGCCAGCAG
    2110      2120      2130      2140      2150      2160
GGCGCTGTACAGGTGGCCAGGGCCAAGGTCGCGGGGCTCAGGTGTCTGCGGGTCGTC
R  S  M  D  V  P  G  R  E  L  P  R  G  S  D  L  R  A  L  L
 A  V  H  G  G  T  R  T  G  A  A  A  G  L  G  S  A  G  A  A
  R  C  T  W  R  D  A  N  W  R  G  G  R  T  W  V  R  W  C  C
```

FIGURE 6 (CONT'D)

```
Q  Q  V  L  D  E  T  C  Y  P  L  G  L  R  C  H  P  A  H  R
 S  R  S  S  T  R  R  V  I  R  S  V  S  D  A  T  R  L  I  A
  A  G  P  R  R  D  V  L  S  A  R  S  P  M  P  P  G  S  S  R
CAGCAGGTCCTCGACGAGACGTGTTATCCGCTCGGTCTCCGATGCCACCCGGCTCATCGC
     2170      2180      2190      2200      2210      2220

 L  L  D  E  V  L  R  T  I  R  E  T  E  S  A  V  R  S  M  A
  A  P  G  R  R  S  T  N  D  A  R  D  G  I  G  G  P  E  D  R
   C  T  R  S  S  V  H  *  G  S  P  R  R  H  W  G  A  *  R  T
GTCGTCCAGGAGCTGCTCTGCACAATAGGCGAGCCAGAGGCTACGGTGGGCCGAGTAGCG

 V  C  D  G  L  G  I  V  P  Y  P  P  L  R  Q  F  R  V  T  T  D
  C  A  T  A  S  G  S  S  P  I  R  C  V  N  S  V  *  P  R  I
   V  R  R  P  R  D  R  P  L  S  A  A  S  I  P  C  N  H  G  S
GTGTGCGACGGCCTCGGGATCGTCCCTATCCGCTGCGTCAATTCCGTGTAACCACCGGAT
     2230      2240      2250      2260      2270      2280

 H  A  V  A  E  P  D  D  G  I  R  Q  T  L  E  T  Y  G  R  I
  T  R  R  G  R  S  R  G  R  D  A  A  D  I  G  H  L  W  P  D
   H  S  P  R  P  I  T  G  *  G  S  R  *  N  R  T  V  V  S  R
CACACGCTGCCGGAGCCCTGGAGCAGGGGGATAGGGACGCAGTTAAGGCACATTGGTGCCTA
```

FIGURE 6 (CONT'D)

```
R  R  K  G  S  S  Q  F  M  T  G  I  G  N  E  L  A  H  T  G
  A  A  R  G  V  R  S  S  *  L  A  S  A  T  N  W  R  T  R  V
    P  Q  G  E  F  A  V  H  D  W  H  R  Q  R  T  G  A  H  G  F
CGGCGCAAGGGGAGTTCGCAGTTCATGACTGGCATCGGCAACGAACTGGCACACGGGT
      2290      2300      2310      2320      2330      2340

A  A  L  P  T  R  L  E  H  S  A  D  A  V  F  Q  R  V  R  T
  G  C  P  S  N  A  T  *  S  Q  C  R  C  R  V  P  A  C  P  N
    R  L  P  L  E  C  N  M  V  P  M  P  L  S  S  A  C  V  P  K
GCGGCGTTCCCCTCAAGGTCAAGTACTGACCGTTGCTTGACCGGTGTGCCCA

F  T  G  L  P  R  R  Q  C  G  S  D  V  V  E  H  P  V  E  R
  S  L  A  C  R  A  D  S  A  A  A  M  W  S  S  I  R  L  S  A
    H  W  P  A  A  P  T  V  R  Q  R  C  G  R  A  S  G  *  A  P
TTCACTGGCCTGCCGCGCCGACAGTGCGGCAGCGATGTGGTCGAGCATCCGGTTGAGCGC
      2350      2360      2370      2380      2390      2400

E  S  A  Q  R  A  S  L  A  A  I  H  D  L  M  R  N  L  A  G
  *  Q  G  A  A  G  V  T  R  C  R  H  P  R  A  D  P  Q  A  G
    V  P  R  G  R  R  C  H  P  L  S  T  T  S  C  G  T  S  R  R
AAGTGACCGGACGCGGCGGGCTGTCACGCGGTCGCCTACACCAGCTCGTAGGCCAACTCGCG
```

FIGURE 6 (CONT'D)

```
R   P   E   L   P   H   L   G   G   G   V   C   V   R   F   G   H   P   D   R
D   P   S   C   P   T   S   V   E   G   F   A   S   G   S   G   T   R   T   G
T   R   V   A   P   P   R   W   R   G   L   R   Q   V   R   A   P   G   P   V
CGACCCGAGTTGCCCCACCTCGGTGGAGGGGGTTGCGTCAGGTTCGGGCACCCGGACCGG
        2410      2420      2430      2440      2450      2460
GCTGGGCTCAACGGGGTGGAGCCACCTCCCAAACGCAGTCCCAAGCCCGTGGGCCTGGCC
S   G   L   Q   G   V   E   T   S   P   N   A   D   P   E   P   V   R   V   P
V   R   T   A   G   G   R   H   L   P   K   R   *   T   R   A   G   P   G   T
G   S   N   G   W   R   P   P   P   P   T   Q   T   L   N   P   C   G   S   R   Y


*   L   D   L   A   A   I   Q   R   *   V   D   D   F   A   R   G   L   R   D
S   L   T   S   P   R   S   N   G   R   S   T   T   S   L   A   V   C   A   T
A   *   P   R   R   D   P   T   V   G   R   R   L   R   S   R   E   A   R   P
TAGCTTGACCTCGCCGCGATCCAACGGTCGAACGGTAGGTCGACGACTTCGCTCGGTTTGCGCGAC
        2470      2480      2490      2500      2510      2520
ATCGAACTGGAGCGGCGCTAGGTTGCCATCCAGCTGCTGAAGGCCAAACGGCGCTG
L   K   V   E   G   R   D   L   P   L   D   V   V   E   S   A   T   Q   A   V
A   Q   G   R   R   S   G   V   T   P   R   R   S   R   E   R   N   A   R   G
S   S   R   A   A   I   W   R   Y   T   S   S   K   A   R   P   K   R   S   R
```

FIGURE 6 (CONT'D)

```
R R N G A S A R L M M T I P A V V A A T
A A T A P A P A * * R F R R S R R P
P Q R R Q R P L D D D S G G R R G D Q
CGCCGCAACGGCGCCAGCGCCCGCTTGATGATGACGATTCCGGCGGTCGTCGCGGCGACC
      2530      2540      2550      2560      2570      2580

A A V A G A G A Q H R N R D D R R G
G C R R W R G S S S S E P P R P S W
R L P A L A R K I I V I G A T T A A V L
GCGGGCGGTTGCCGCGGTCGCGGGGCGAACTACTGCTAAGGCCGCCAGCAGCCGCCGCTGG

N A I T V T I P K M I S I C N I V A S T
T Q S P * R F R K * S A S A T S W R R R
R N H R D D S E N D Q H L Q H R G V D V
AACGCAATCACCGTGACGATTCCGAAAATGATCAGCATCTGCAACATCGTGGCGTCGACG
      2590      2600      2610      2620      2630      2640

V C D G H R N R E H D A D A V D H R R R
R L * R S S E S F S * C R C C R P T S T
A I V T V I G F I L M Q L M T A D V N
TTGCCGTTAGTGGCACTGCTAAGGCTTTTACTAGTCGTAGACGTTGTAGCACCGCAGCTGC

L P I D R P V T M T S C P F R L G A A S
C P S T G R * R * R R A R E G S E R P A
A H R Q A G D D D V V P V S A R S G Q H
TTGCCCATCGACAGGCCGGTGACGATGACGTCGTGCCCGTTCGGCTCGGAGGCGGCCAGC
```

FIGURE 7

```
  1 GAA TTC CAA CCG TCG GTG CAG ATC CAG GTC TAT CAG GGG GAG CGT GAG  48
    Glu Phe Gln Pro Ser Val Gln Ile Gln Val Tyr Gln Gly Glu Arg Glu

 49 ATC GCC GCG CAC AAC AAG TTG CTC GGG TCC TTC GAG CTG ACC GGC ATC  96
    Ile Ala Ala His Asn Lys Leu Leu Gly Ser Phe Glu Leu Thr Gly Ile

 97 CCG CCG GCG CCG CGG GGG ATT CCG CAG ATC GAG GTC ACT TTC GAC ATC 144
    Pro Pro Ala Pro Arg Gly Ile Pro Gln Ile Glu Val Thr Phe Asp Ile

145 GAC GCC AAC GGC ATT GTG CAC GTC ACC GCC AAG GAC AAG GGC ACC GGC 192
    Asp Ala Asn Gly Ile Val His Val Thr Ala Lys Asp Lys Gly Thr Gly
```

FIGURE 7 (CONT'D)

```
93  AAG GAG AAC ACG ATC CGA ATC CAG GAA GGC TCG GGC CTG TCC AAG GAA  240
    Lys Glu Asn Thr Ile Arg Ile Gln Glu Gly Ser Gly Leu Ser Lys Glu

11  GAC ATT GAC CGC ATG ATC AAG GAC GCC GAA GCG CAC GCC GAG GAG GAT  288
    Asp Ile Asp Arg MET Ile Lys Asp Ala Glu Ala His Ala Glu Glu Asp

89  CGC AAG CGT CGC GAG GAG GCC GAT GTT CGT AAT CAA GCC GAG ACA TTG  336
    Arg Lys Arg Arg Glu Glu Ala Asp Val Arg Asn Gln Ala Glu Thr Leu

 7  GTC TAC CAG ACG GAG AAG TTC GTC AAA GAA CAG CGT GAG GCC GAG GGT  384
    Val Tyr Gln Thr Glu Lys Phe Val Lys Glu Gln Arg Glu Ala Glu Gly
```

FIGURE 7 (CONT'D)

```
385  GGT TCG AAG GTA CCT GAA GAC ACG CTG AAC AAG GTT GAT GCC GCG GTG  432
     Gly Ser Lys Val Pro Glu Asp Thr Leu Asn Lys Val Asp Ala Ala Val

133  GCG GAA GCG GAA GGC GGC ACT TGG CGG ATC GGA TAT TTC GGC CAT CAA  480
     Ala Glu Ala Glu Gly Gly Thr Trp Arg Ile Gly Tyr Phe Gly His Gln

181  GTC GGC GAT GGA GAA GCT GGG CCA GGA GTC GCA GGC TCT GGG GCA AGC  528
     Val Gly Asp Gly Glu Ala Gly Pro Gly Val Ala Gly Ser Gly Ala Ser

29   GAT CTA CGA AGC AGC TCA GGC TGC GTC ACA GGC CAC TGG CGC TGC CCA  576
     Asp Leu Arg Ser Ser Ser Gly Cys Val Thr Gly His Trp Arg Cys Pro

77   CCC CGG CGG CGA GCC GGG CGG TGC CCA CCC CGG CTC GGC  615
     Pro Arg Arg Arg Ala Gly Arg Cys Pro Pro Arg Leu Gly
```

FIGURE 8

```
5' TCGAACGAGGGGGGGGTGACCCGGTGCGGGGCTTCTTGCACTCGGCCATAGGCGGAGTGCTAAG
3' AGCTTGCTCCCCGCACTGGGCCACGCCCCGAAGAACGTGAGCCGGTATCCGCTCACGATTC
                10        20        30        40        50        60

AATAACGGTTGGCCACTCGGCCGACCGGTCGAGTGCTAGGTCGGGACGGTCGGGTGAGGCCAGGCCCGTC
TTATTGCCAACGGTGGCCGTGGCCACTGGCCAGCTCACGATCCAGCCCTGCCACTCCGGTCCGGGCCAG
             70        80        90        100       110       120

GTCGCAGCCAGGAGGACAACTTGAGCGTCCGTCGCGGGCCACTGCCGCCCCGCC
CAGCCGTCGGCTCACCGTCGCCTCCTGTTGAACTTGCGAGGCGGGCCAGGCGGGGC.CGG
            130       140       150       160       170       180
                      R   G   C   R   H   P   V   T   P   V   S   S   P   T   R   R
AGGGTAAGTAGGCGGGGTTGCCGTCAGGGTGACCCCGGTTCATCCCGATCCGGAGGA
TCGCATTCATCGCCCCAACGGCCAGTGGCCACTGGGGGCAAAGTAGGGGCTAGGGGCTCCT
            190       200       210       220       230       240
  N   H   F   A   N   A   K   T   I   A   Y   D   E   A   R   R   G   L   E
ATCACTTCGGCAATGGCCAAGACAATTGCGTAGCCAACGCCCGTCGCCCGGGCCTCGAGC
TAGTGAAGCCGTTACCGGTTCTGTTAACGGATCGCTTCGGGCAGCGGCCGGAGCTCG
            250       260       270       280       290       300
  R   G   L   N   A   L   A   D   A   V   K   V   T   L   G   P   K   G   R   N
GGGGCTTGAACGCCCTCGCCCATGCGGTAAAGGTGACATTGGGCCCCCAAGGGCCCAACG
CCCCGAACTTGCGGGAGCGGGTACGGCCATTCCACTGTAACCCGGGGTTCCCGGGGTTGC
            310       320       330       340       350       360
  V   V   L   E   K   K   W   G   A   P   T   I   T   N   D   G   V   S   I   A
TCGTCCTGGAAAGAAGTGGGGTGCCCCACGTGGGGTGCCCCACGATCACCAACGATGGTGTCCATCGGCCA
AGCAGGACCTTTCTTCACCCCCACGGGGGTGCCTACCACACAGGTAGCGGGT
            370       380       390       400       410       420
```

FIGURE 8 (CONT'D)

```
  K   E   I   E   L   E   D   P   Y   E   K   I   G   A   E   L   V   K   E   V
AGGAGATCGAGCTGGAGGATCCGTACCGAAGAGATCGGCGCCCGAGCTGGTCAAGAGGTAG
TCCTCTAGCTCGACCTCCTAGGCATGGCTTCTCTAGCCGCGGGCTCGACCAGTTCTCCATC
                    430                 440                 450                 460                 470                 480

  A   K   K   T   D   D   V   A   G   D   G   T   T   T   A   T   V   L   A   Q
CCAAGAAGACCGATGACGTCGCCGGTGACGGCACCACGACGGCCACCGTGCTGGCCCAGG
GGTTCTTCTGGCTACTGCAGCGGCCACTGCCGGTGGTGCTGCCGGTGGCACGACGGGTCC
                    490                 500                 510                 520                 530                 540

  A   L   V   R   E   G   L   R   N   V   A   G   A   N   P   L   G   L   K
CGTTGGTTCCGCGAGGGCCTGCGCAACGTCGCGGCCGCCAACCCGGTCGGTCTCAAAC
GCAACCAAGCCCTCCCGGACGCGTTGCAGCGCCGGCGGTTGGGCCAGCCAGAGTTTG
                    550                 560                 570                 580                 590                 600

  R   G   I   E   K   A   V   E   K   V   T   E   T   L   L   K   G   A   K   E
GCGGCATCGAAAAGGCCGTGGAAGAAGGTCACCGAGACCCTGCTCAAGGGCGCCAAGGAGG
CGCCGTAGCTTTTCCGGCACCTTCTTCCAGTGGCTCTGGGACGAGTTCCCGCGGTTCCTCC
                    610                 620                 630                 640                 650                 660

  V   E   T   K   E   Q   I   A   A   T   A   A   I   S   A   G   D   Q   S   I
TCGAGACCAAGGAGCAGATTGCGGCCACCGCAGCGATTTCGGCGGGTGACCAGTCCATCG
AGCTCTGGTTCCTCGTCTAAACGCCGGTGGCGTCGCTAAAGCCGCCCACTGGTCAGGTAGC
                    670                 680                 690                 700                 710                 720

  G   D   L   I   A   E   A   M   D   K   V   G   N   E   G   V   I   T   V   E
GTGACCTGATCGCCGAGGCGATGGACAAGGTGGGCAACGAGGGCGTCATCACCGTCGAGG
CACTGGACTAGCGGCTCCGCTACCTGTTCCACCCGTTGCTCCCGCAGTAGTGGCAGCTCC
                    730                 740                 750                 760                 770                 780
```

FIGURE 8 (CONT'D)

```
E  S  N  T  F  G  L  L  E  L  T  E  G  M  R  F  D  K  G
AGTCCAACACCTTTGGGCTGCAGCTCGAGCTCCACCGAGGGTATGCGGTTCGACAAGGGCT
TCAGGTTGTGGAAACCCGACGTCGAGCTCGAGGTGGCTCCCATACGCCAAGCTGTTCCCGA
          790        800        810        820        830        840

Y  I  S  G  Y  F  V  T  D  P  E  R  Q  E  A  V  L  E  D  P
ACATCTCGGGGTACTTCGTGACCGACCCGGAGAGGCAGGAGGCGGTCCTTGGAGGACCCCT
TGTAGAGCCCCATGAAGCACTGGCTGGGCCTCGCCAGTCCTCCGCCAGGAACCTCCTGGGGA
          850        860        870        880        890        900

Y  L  L  V  S  S  K  V  S  T  V  K  D  L  P  L  L  E
ACATCCTGCTGGTCAGCTCCAAGGTGTCCACTGTCAAGGATCTGCCGCTGCTCGAGA
TGTAGGACGACCAGTCGAGGTTCCACAGGTGACAGTTCCTAGACGGCGACGAGCTCT
          910        920        930        940        950        960

K  V  I  G  A  G  K  P  L  L  I  I  A  E  D  V  E  G  E  A
AGGTCATCGGAGCCGGTAAGCCGCTGCTGATCATCGCCGAGGACGTCGAGGGCGAGGCCGC
TCCAGTAGCCTCGGCCATTCGGCGACGACTAGTAGCGGCTCCTGCAGCTCCCGCTCCGGCG
          970        980        990        1000       1010       1020

L  S  T  L  V  V  N  K  I  R  G  T  F  K  S  V  A  V  K  A
TGTCCACCCTGGTCGTCAACAAGATCCGCGGCACCTTCAAGTCGGTGGCCGTCAAGGCTC
ACAGGTGGGACCAGCAGTTGTTCTAGGCGCCGTGGAAGTTCAGCCACCGGCAGTTCCGAG
          1030       1040       1050       1060       1070       1080

P  G  F  G  D  R  R  K  A  M  L  Q  D  M  A  I  L  T  G  G
CCGGCTTCGGCGACCGCCGCAAGGCCATGCTGCAGGATATGGCCATTCTCACCGGTGGTC
GGCCGAAGCCGCTGGCGGCGTTCCGGTACGACGTCCTATACCGGTAAGAGTGGCCACCAG
          1090       1100       1110       1120       1130       1140

Q  V  I  S  E  V  G  L  T  L  E  N  A  D  L  S  L  G
AGGTGATCAGCGAAGAGGTCGGCCTGACCCTGGAGAACGCCGACCTGTCGCTAGGCCA
TCCACTAGTCGCTTCTCCAGCCGGACTGGGACCTCTTGCGGCTGGACAGCGATCCGGT
          1150       1160       1170       1180       1190       1200
```

FIGURE 8 (CONT'D)

```
      K   A   R   K   V   V   T   K   D   E   T   T   I   V   E   G   A   G   D
AGGCCCGCCAAGGTCGTGGTCACCAAGGACGAGACCACCATCGTCGAGGGCGCCGGTGACA
TCCGGGCGGTTCCAGCACCAGTGGTTCCTGCTCTGGTGGTAGCAGCTCCCGCGGCCACTGT
            1210            1220            1230            1240            1250            1260

      T   D   A   I   A   G   R   V   A   Q   I   R   Q   E   I   E   N   S   D   S
CCGACGGCCATCGCCGGACGAGTGGCCCAGATCCGCCAGGAGATCGAGAACAGGACTCCG
GGCTGCCGGTAGCGGCCTGCTCACCGGGTCTAGGCGGTCCTCTAGCTCTTGTCGTGAGGC
            1270            1280            1290            1300            1310            1320

      D   Y   D   R   E   K   L   Q   E   R   L   A   K   L   A   G   G   V   A   V
ACTACGACCGTGAGAAGCTGCAGGAGCGGCTGGCCAAGCTGGCCGGTGGTGTCGCGGGTGA
TGATGCTGGCACTCTTCGACGTCCTCGCCGACCGGTTCGACCGGCCACCACAGCGCCCACT
            1330            1340            1350            1360            1370            1380

      I   K   G   A   A   T   E   V   E   L   K   E   K   H   R   I   E   D
TCAAGGGCGCCGCCACCGAGGTCGAACTCAAGGAGAAGCACCGCATCGAGGATG
AGTTCCCGCGGCGGTGGCTCCAGCTTGAGTTCCTCTTCGTGGCGTAGCTCCTAC
            1390            1400            1410            1420            1430            1440

      A   V   R   N   A   K   A   V   E   E   G   I   V   A   G   G   V   T
CGGTTCGCAATGCCAAGGCtGCCGTCGAGGAGGGCATCGTCGCCGGTGGTGTGACGC
GCCAAGCGTTACGGTTCCGGCGGCAGCTCCTCCCGTAGCAGCGGCCACCACACTGCCG
            1450            1460            1470            1480            1490            1500

      L   L   Q   A   A   P   T   L   D   E   L   K   L   E   G   D   E   A   T   G
TGTTGCAAGcGGCCCCGACCCTGGACGAGCTGAAGCTCGAGGGCGACGAGGCCACCGGGCG
ACAACGTTCGCCGGGGCTGGGACCTGCTCGACTTCGAGCTCCCGCTGCTCCGGTGGCCCGC
            1510            1520            1530            1540            1550            1560

      A   N   I   V   K   V   A   L   E   A   P   L   K   Q   I   A   F   N   S   G
CCAACATCGTGAAGGTGGCCCTGGAGGCCCCGCTGAAGCAGATCGCCTTCAACTCCGGGC
GGTTGTAGCACTTCCACCGGGACCTCCGGGGCGACTTCGTCTAGCGGAAGTTGAGGCCCG
            1570            1580            1590            1600            1610            1620
```

FIGURE 8 (CONT'D)

```
L   E   P   G   V   A   E   K   V   R   N   L   P   A   G   H   G   L   N
TGGAGCCGGGCTGGTGGCCGAGAAGGTGGCCAACCTGCCGGCTGGCCACGGACTGAACG
ACCTCGGCCCGACCACCGGCTCTTCCACCGGCGTTGGACGGCCGACCGGTGCCTGACTTGC
    1630      1640      1650      1660      1670      1680

A   Q   T   G   V   Y   E   D   L   L   A   G   V   A   D   P   V   K   V
CTCAGACCCGGTGTCTACGAGGATCTGCTCGCTGGCGTGGCCGGTTGCTGACCCGGTCAAGGTGA
GAGTCTGGGCCACAGATGCTCCTAGACGACGAGCGACCGCCAACGACTGGGCCAGTTCCACT
    1690      1700      1710      1720      1730      1740

T   R   S   A   L   Q   N   A   A   S   I   A   G   L   F   L   T   E   A
CCCGTTCGGGCTGCAGAATGCGGGCGTCCATCGGCGGGGCTGTTCCTGACCACCGAGGCCG
GGGCAAGCCCGACGTCTTACGCCCGCAGGTAGCCGCCCCGACAAGGACTGGTGGCTCCCGGC
    1750      1760      1770      1780      1790      1800

V   V   A   D   K   P   E   K   E   K   A   S   V   P   G   G   G   D   M   G
TCGTTGCCGACAAGCCCGAAAAGGAGAAGAAGGCTTCCGTTCCCGGTGGCGGCGGCGACATGGGGTG
AGCAACGGGCTGTTCGGGCCTTTCCTCTTCTTCCGAAGGCAAGGGCCACCGCCGCCGCTGTACCCAC
    1810      1820      1830      1840      1850      1880

G   M   D   F
GCATGGATTTCTGACCCCGGGCCCCGGACAAGTCCGACGAGCCCGGTCCCCTTTGTGGGGCC
CGTACCTAAAGACTGGGGCCCGGGGCCTGTTCAGGCTGCTCGGGCAGGGGAAACACCCCGG
    1870      1880      1890      1900      1910      1920

GGGCTCCTCTGGTTGGGAGCTACGGTACCGGAGAACCACCAGCAGTCGTGTAGGCAACCTT
CCGAGGAGACCAACCCGGTCGGTGGCCATGGCCACATCCGGTTGGAA
    1930      1940      1950      1960      1970      1980

TGGCCGCTGTGTGGGCCGAGTCGGGGGCGCCCGTCTCGGTGCAGCAGCGGCGGGATGGGTACGA
ACCGGCGACACCCCGGCTCAGCCCCCGGCGGCCAGAGCCACGTCGCGCGGCGCTACCCATGCT
    1990      2000      2010      2020      2030      2040
```

FIGURE 8 (CONT'D)

```
CACCCCAGCCGCGGCCGGTGTCGTCATCGGGGGCCTGCTCCGACGCGTCCTGGGCCACGGCCGTCGA
GTGGGCGTCGCCGCCACAGCCAGTAGCCCCGGACGCAGCCCGGTGCCGGGACCCGTGCCGGGCAGCT
         2050          2060          2070          2080          2090          2100

CGATCCAGCCGGAGTAGCCGGCTAGGATCCGGATCCGGATCGGGATGGGCGGCCACAACAGGGTGACTTCGCTGCGGT
GCTAGGTCGCTCATCGGGGCGATCCTAGCCTACCGCCGGGTGTTGTCCCACTGAAGCGACGGCCA
         2110          2120          2130          2140          2150          2160

GGGCCAGGTTTTGCCGCGGTACGACCCCCGATCAGGCCGACGTCGACCACTGCCCCGGGGTC
CCCGGTCCAAAACGGGCATGCTGGGGGCTGGGGGCTGGCTGCAGCTGGGTGACGGGCCCCAG
         2170          2180          2190          2200          2210          2220

CATCGGGGGCCGTCGGGGAGTTCGGCGCAGCACCGGGCTCGACTGCCACCGTGTGCACGCGGAT
GTAGCCCCGGCAGCCGTCAAGCCGCGTCGTGGCCGAGCTGACGGTGGCACACGTGCGCCTA
         2230          2240          2250          2260          2270          2280

GGCCATCATCGACGGTGATCAGGTAAGCGGAACGGGTAGTCGGGCAAGGCCGGGCGGCCAGCC
CCGGTAGTCGCTGCCACTAGTCCATTCGCTTGCCCATCGAGCCCGTTCCGCCGCCGGGTCGG
         2290          2300          2310          2320          2330          2340

GTTTGAGGGTCTACCTTTTGGCACCCACCGGATTCGGAGGATAGGCGCCCGATGTGTTACTC
CAAACTCCAGATGGGAAAACCGTGGGGTGCCTAGGGTGCCCTATCCGCGGGCTACACAATGAG
         2350          2360          2370          2380          2390          2400

CGAACCGACCGGCTGCCCGATCCGCCGGGGATCCCGCGGGCTGGGCGTAGGCCGGATTCGGGGGCTCGG
GCTTGGCTGGCCGACGGGCTAGGGCGGCTAGGCGCCGACGGGCCCATCCGCCCTAAGCCGCCCCCGAGCC
         2410          2420          2430          2440          2450          2460
              S  G  V  P  Q  G  I  R  P  S  A  Y  A  S  E  R  D  P  S  P
                                                                           *    E
GTAGAAGTTCGACTTGGGGATGCCCAGCCCGGGGGTACTCGGCCTCACGCACGGGTATT
CATCTTCAAGCTGAACCCCTACGGGTCGGCCCCCATGAGCCGGAGTGCGTGCGCCCATAA
         2470          2480          2490          2500          2510          2520
    Y  F  N  S  K  P  I  G  S  G  P  I  S  P  E  R  V  A  T  H
```

FIGURE 8 (CONT'D)

```
           2530      2540      2550      2560      2570      2580
CCGGCAAGCCCGGAGTCGGCTGCTGCCCGAGTTGACGAAGCTCGGGTAGCTGGTGGTGCCAGGGCT
GGCGTTCGGGCTCAGCAACGACGGGCTCAACTGCTTCGAGCCATCGGACCACGGTCCCGA
 R   L   G   S   D   H   S   G   S   H   V   F   S   P   Y   S   T   G   P   S

           2590      2600      2610      2620      2630      2640
TCTAAGGCCCGGGGTTTGCCCCCAGCCGGGGGGCTCGGTGGCTCCGGGGGGGTTCGGGTT
AGATTCCGGGCCCAAACGCGGGGTCGGGGCTCGGTGTCCTCGTGACCGGCGGTGGCCAAGCCCAA
 R   L   G   P   H   A   G   S   G   A   A   A   S   G   S   G   P   H   P   H

           2650      2660      2670      2680      2690      2700
GCCTGAGTCCAGGCCCGCCCAACAGGAGCCACTGGCCCGGGGGGCCGGGCGTGTTGGTCAG
CGGACTCAGGTCCGGGCGGTTGTCCTCGGTGACCGGCGCCCGCCCGCCGCCCGCCACAACCAGTC
 G   S   D   L   G   G   V   P   A   S   A   P   A   A   V   P   T   H   T   L

           2710      2720      2730      2740      2750      2760
GCCCGAGTTGAGGACGTTCGGCCAGGCCCGTGTTGGACACCCGCCGTTGATCCGAGGGGCGGA
CGGGCTCAACTCCTGCAAGCGGGTCCGGGCACAACCTCTGGCGGGCAACTAGGCTCCCGCCT
 G   S   H   L   V   H   A   L   G   H   Q   L   G   G   T   S   G   L   A   S

           2770      2780      2790      2800      2810      2820
GGGGAGGATGCCCCAACTCAAAGCCCGGCTTCATCGCCCGGTAGCCGGCGGA
CCGGCTCCTACGGCTTGAGTTCGGGCGGCTTTCGGGCGGCACGGTACGGCGGCACGGCGGCCGCCT
 A   L   I   G   S   S   L   A   A   T   S   M   G   G   T   A   Y   G   A   S

           2830      2840      2850      2860      2870      2880
GCTGACCAAGGCCGGGGCCCTCCCAGCCAGCCGGCCTTCGCCCGGCGTTTGCATCCCCGC
CGACTGGGTTCCGGGGCCGGAGGTCGGCGGCGGAGGATTCCGGCCGGCCAAAACGTAGGGGGCG
 S   V   L   A   E   S   G   A   A   S   G   L   A   A   H   Q   M   G   A

           2890      2900      2910      2920      2930      2940
GTTCCAGAAGCTGGTGTTGAGGCTGCCGCTGCCGGAGGCCCGGTTGATTGTCCCCGA
CAAGGTCTTCGACCACAACTCCGACGGCGACTGCCGGGCGGCCAACTAACAGGGGGCT
 H   W   F   S   T   H   L   S   G   A   S   G   L   G   A   H   I   T   G   S
```

FIGURE 8 (CONT'D)

```
          2960      2970      2980      2990      3000
GGTCCCGATGCCGGCCTGTTGTTCAGGGGAGCCCGAATTCCCGATGCCCGATGTTCCGGCTGCCGGA
CCAGGGCTACGGCCGGGACAAGTCCCTCGGGGCTTAAGGGCTACGGGCTACAAAGGGCGACGGGCCT
    T    I    G    S    H    L    S    G    S    N    G    I    G    I    N    G    S    G    S

       3010      3020      3030      3040      3050      3060
GTTGAATAAGCCGACGTTGCCGGTGCCCGCCCAGTTCCCGAAGCCGATGTTGCCGTACCCGA
CAACTTATTCGGCTGCAACGGCCACGGGCGGGTCAAGGGCTTCGGCTACAACGGGATGGGCT
    N    F    L    G    V    N    G    T    G    S    N    G    F    G    I    N    G    S    G    S

       3070      3080      3090      3100      3110      3120
GTTGAAGCCGCCCGAAAACCCATCTGGTGATCACCGGGTGATCCCGAACCCGATATTCCCGCT
CAACTTCGGCGGGCTTTTGGGGTAGACCACTAGTGGCCACTAGGGCTTGGGCTATAAGGGCGA
    N    F    G    G    F    G    M    Q    H    D    G    T    I    G    F    G    I    N    G    S

       3130      3140      3150      3160      3170      3180
ACCGGTGTTGCCCGAAGCCGATATTCCCGTCGGCCCGAGGTTGCCGGAGGCCCAGGTTGCCGCT
TGGCCACAACGGGCTTCGGCTATAAGGGCAGCCGGGCTCCAACGGGTCCCGGGTCCAACGGGCCA
    G    T    N    G    F    G    I    N    G    D    G    L    G    L    G    L    N    G    S

       3190      3200      3210      3220      3230      3240
GCCGGTGTTGCCCGATGTTGCCCGGTGTTGCCGGTGTTGCCGGATGTTGTTGTT
CGGCCACAACGGGCTACAACGGGCTACAACGGGCTACAACGGGCTACAACAACAA
    G    T    N    G    S    G    I    N    G    T    G    T    N    G    S    G    I    N    N    N

       3250      3260      3270      3280      3290      3300
GCCGATGTTGTTGCCGGATGTTGTTGCCGGTTGTTGCCGGTGTTGCCGAA
CGGCTACAACAACGGGCTACAACAACGGGCTACAACAACGGGCCACAACGGGCTT
    G    I    N    N    N    G    I    N    N    N    G    I    N    G    S    G    T    N    G    F

       3310      3320      3330      3340      3350      3360
GCCCAGATTGATCTCGGCCGTTCTTGCCGGATGTCGGATGTCCGGAGGTTCCGCAAGACCTGCTG
CGGGTCTAACTAGACCGGCAAGAACGGGCTACAGCCTACAGGCCTCCAAGGCGTTCTGGACGAC
    G    L    H    I    Q    G    N    K    G    I    D    I    G    L    H    R    L    V    Q    Q
```

FIGURE 8 (CONT'D)

```
      3370      3380      3390      3400      3410      3420
CCAGGGGGCGCCAGTTGTGCCGACGGGGCCCGCAGACACGGCCATCAACGAAGTGGTAACCAGCCATCGCCGC
GGTCCCGCGGGTCAACACGGCTGCCCGGGCGTCTGTGCCGGTAGTTGCTTCACCATTGGTCGGTCGGTAGGCGGCG
  W   P   A   L   Q   A   V   A   S   A   D   F   H   Y   G   A   M   A   A

      3430      3440      3450      3460      3470      3480
CACGTCCAATGCCCACATTTGCTCGTATGCCCGCTCGACGTTCCATGAGCGCCGGAGCCGTT
GTGCCAGGTTACGGGTGTAAACGAGCATACGGGCGAGCTGCAAGGTACTCGCGGCCTCGCAA
  V   D   L   A   W   M   Q   E   Y   A   A   E   V   D   M   L   A   P   A   H

      3490      3500      3510      3520      3530      3540
CTGCCCAAACCAGTTCGTAGCTGCCAGCAGCTGCATCAGGCCACGATTGGCCGCTACCAC
GACGGGTTTGGTCAAGCATCGACGGTCGTCGACGTAGTCCGGTGCTAACGGCGATGGTG
  Q   G   F   W   N   T   A   A   L   L   Q   M   L   G   R   N   A   A   V   V

      3550      3560      3570      3580      3590      3600
TGCCGGCTGCACGGTGGCCGCCAGGCCCTCGAACGCGGTCGGTTGCCATGGCCCTG
ACGGCCGACGTGCCACCGGCGGTCGGGGAGCTTGCGCCAGGACAACGGTACCGGAC
  A   P   Q   V   T   A   A   L   A   A   E   F   A   T   A   T   A   M   A   Q

      3610      3620      3630      3640      3650      3660
TGCGGGCCGCTTGTTCCCGCCGCTCCCCGTGCTGAGCCAGGCTAGGTACTGGGTTGC
ACGCCCGGCGAACAAGGGCGGCGACGGGCGACGGGGCACGGTCCGATCCATGACCAACG
  A   A   A   Q   E   A   Q   A   A   T   S   L   W   A   L   Y   Q   T   A

      3670      3680      3690      3700      3710      3720
GACGGGCCATCATCGCCGCGCGGACCCGGGGACCCAGGCCAGCCCACTAGTTCAGTTCCGATGT
CTGCCGGTAGTAGCGGCGCGCCCTGGGTCGGTCGGGTGATCAGTCAGTTCAAGTTCAAGCCTACA
  V   A   M   M   A   A   A   S   P   G   L   W   A   G   S   T   L   E   S   T

      3730      3740      3750      3760      3770      3780
GACGGAGCCCAAGGCGACGACCTATTGACGCGGACCAATTCTTCGGCCGCCAGGCGGT
CTGCCCTCGGGTTCCGCTGCTGGATAACGTCGCGCCTGGTTAAGAAGCCGGCGGGGTCCGCCA
  V   S   G   L   S   A   I   S   A   L   E   E   A   L   E   G   W   A   T
```

FIGURE 8 (CONT'D)

```
         3790       3800       3810       3820       3830       3840
GGCCCGCAGCAATTAGCCGCGGGTCCCGACCCGGACCGGCAAACATCAGTGCCGGAATTGATCTC
CCGGCCGTCCTTAATCGCCACGGCTGGGCCCTGGCCGGTTGTAGTCGCGGTTAACTAGAG
 A  A  A  I  L  P  G  S  G  P  G  A  F  M  L  A  S  N  I  E
         3850       3860       3870       3880       3890       3900
TGGCGGCAACCACCACGCAAAATGCGGGGTTGTCAGCCGATCCAACTTAACTGTCAGGGACCG
ACCCGCCGTTGGCGTGCGTTTTACGCCCCAACAGTCGGCTAGGTTGAATTGACAGTCGGTGGC
 P  P  L  W  A  F  H  P  S  T  L  R  D  L  K  V  T  L  S  R
         3910       3920       3930       3940       3950       3960
TTGCCGTGGCGGTATCGGCACTTCAATACCACTCATCTTTGGGGTCATCTTTGGAGCGCC
AACGGCACCGCCATAGCCGTGAAGTTATGGGTGAGTAGAAACCCAGTAGAAACCTCGCGG
 Q  R  P  P  I  P  V  E  I  G  S  M  K  P  T  M  K  P  A  G
         3970       3980       3990       4000       4010       4020
CCTAGGAACCGCCGCCAGCTTACCTAGTCCCGGGTAGGGCCCGACTGGCGGCCGGGATGCAGC
GGATCCTTGCGGCGGTCGAATGGATCAGGCGCCCATCCCGGGCTGACCGCCGGCCCTACGTCG
 R  P  V  A  L  K  G  L  G  P  Y  P  G  V  P  P  R  S  A  A
         4030       4040       4050       4060       4070       4080
TGAGGGTCTGCCACCTGCCCCGTAATGTCGCTGGTATGGCAAGCACCGAAGGCACGGCGGGCCC
ACTCCCAGACGGGTGACGGGGCATTACCAGCGACCATACCGTTCGTGGCTGCGGCGCCGCGGG
 S  P  R  G  G  A  G  Y  H  R  Q  Y  P  L  C  R  R  R  P  G
         4090       4100       4110       4120       4130       4140
AAGAGTTGCTCCGGACGTCCCGGACGGCGTTCACCCGGTTGATCGGACAACATGTCGACTCACCGACG
TTCTCAACGAGGCCTGCAGGGCCTGCCAAGTGGGCCAACTAGCTTGTACAGCTGTTGAGTGGCTGC.
 L  T  A  G  R  R  T  *
         4150       4160       4170       4180       4190       4200
GCCTCACCGACCAACTCGCCTGCTACCCCGCCCCAGCCCCAACAGCAGGCCAACAGCATTGCCTGGC
CGGAGTGGCTGGGTTGAGCGGACGACGAGATGGGGGGCGGTCGGGGTCGGGGTGTCGTAACGGCACCG
```

FIGURE 8 (CONT'D)

```
     4210      4220      4230      4240      4250      4260
TGCTCTGGCACAGCCGCCCGGGTGCCAGGATATACACGGTCGCCCATGTGGCCGGGCGTGGAAG
ACGAGACCGGTCGCGGCGGGCCCACGTCCTATATGTCCAGCGGGTACACCGGCCGGCACCTTC

     4270      4280      4290      4300      4310      4320
AGGTGTGGACCCCGGACGGTTGGGGTGGACCGGCTTTGGGTTAGATCTGCCGGGCCACGACA
TCCACACCTGGGCGCTGCCAACCCACCTGGGCGAAACCCAATCTAGACGGGCCGGCCGTGCTGT

     4330      4340      4350      4360      4370
CCGGATATGGACACCGTCCCGAGGATGTGGCCGAAGGTACGGGCCACCCGCCCGACGGAATTC   3'
GGCCTATACCTGTGGCCAGGGGCTCCTACACCGGCTTCCATGCCCGTGGGGCGGCCTGCCTTAAG   5'
```